



TECHNISCHE
UNIVERSITÄT
DARMSTADT

ULB

Datenbasierte Subgridskalen-Modellierung reaktiver Konzentrationsgrenzschichten an freiaufsteigenden Einzelblasen

Kiefer, Alexander
(2020)

DOI (TUprints): <https://doi.org/10.25534/tuprints-00011667>

Lizenz:



CC-BY 4.0 International - Creative Commons, Attribution

Publikationstyp: Master Thesis

Fachbereich: 16 Department of Mechanical Engineering

Quelle des Originals: <https://tuprints.ulb.tu-darmstadt.de/11667>

Datenbasierte Subgridskalen - Modellierung reaktiver Konzentrationsgrenzschichten an freiaufsteigenden Einzelblasen

Data-driven subgrid-scale modeling of reactive concentration boundary layers at single rising bubbles

Master-Thesis von Alexander Kiefer

Tag der Einreichung:

1. Gutachten: Prof. Dr. Dieter Bothe
2. Gutachten: Prof. Dr.-Ing. Peter Stephan



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Mathematik
Mathematical Modeling and Analysis

Datenbasierte Subgridskalen - Modellierung reaktiver Konzentrationsgrenzschichten an freiaufsteigenden Einzelblasen
Data-driven subgrid-scale modeling of reactive concentration boundary layers at single rising bubbles

Vorgelegte Master-Thesis von Alexander Kiefer

1. Gutachten: Prof. Dr. Dieter Bothe
2. Gutachten: Prof. Dr.-Ing. Peter Stephan

Tag der Einreichung:

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-116673

URL: <http://tuprints.ulb.tu-darmstadt.de/11667>

Dieses Dokument wird bereitgestellt von tuprints,
E-Publishing-Service der TU Darmstadt
<http://tuprints.ulb.tu-darmstadt.de>
tuprints@ulb.tu-darmstadt.de



Die Veröffentlichung steht unter folgender Creative Commons Lizenz:
Namensnennung 4.0 International
<http://creativecommons.org/licenses/by/4.0/>

Erklärung zur Abschlussarbeit gemäß § 22 Abs. 7 APB TU Darmstadt

Hiermit versichere ich, **Alexander Kiefer**, die vorliegende Master-Thesis gemäß § 22 Abs. 7 APB der TU Darmstadt ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Falle eines Plagiats (§38 Abs.2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Datum :

01.04.2020

Unterschrift:

Alexander Kiefer

Kurzzusammenfassung

Betrachtet wird das Problem großer Schmidt-Zahlen an einzelnen, freiaufsteigenden Gasblasen mit einer zusätzlichen chemischen Reaktion der Transferspezies. In so einem Fall ist die Konzentrationsgrenzschicht deutlich kleiner als die Strömungsgrenzschicht. Für die Direkte Numerische Simulation bedeutet dies die Verwendung von sehr feinen Gittern, welche den Simulationsaufwand erhöhen. Um dem zu begegnen, verwendeten vorhergehende Arbeiten ein Subgridskalen-Modell für den Speziestransport, womit gröbere Gitter benutzt werden können. Der ungenau wiedergegebene Stofftransport wurde durch Eingriffe des Modells korrigiert. Dies geschieht durch Anpassung der Fluss- und Quellterme der Speziestransportgleichungen in der Simulation. Das Subgridskalen-Modell wurde durch einen datenbasierten Ansatz mit Hilfe eines neuronalen Netzes berechnet. Die dafür notwendigen Trainingsdaten wurden durch Simulationen eines vereinfachten Ersatzproblems gewonnen. Aufbauend auf dieser Vorgehensweise wird in dieser Arbeit ebenfalls ein datenbasiertes Subgridskalen-Modell verwendet. Der Prozess der Trainingsdatengenerierung wird weiter vereinfacht. Es wird ein eindimensionales Modell abgeleitet, mit dem die Trainingsdaten berechnet werden. Dies reduziert den Aufwand für die Datengenerierung. Anstatt einer Zerfallsreaktion für die Transferspezies wird in dieser Arbeit die Reaktion der Transferspezies mit einer, in der Flüssigkeit gelösten, Spezies betrachtet, somit treten zwei zusätzliche Spezies auf. Um den Prozess des Netztrainings zu erleichtern wird eine Reduktion des Trainingsdatensatzes durchgeführt, um redundante Daten herauszufiltern. In dieser Arbeit werden zwei Reduktionsalgorithmen entwickelt, miteinander verglichen und schließlich kombiniert. Erste Tests des Subgridskalen-Modells an einer einfachen Blasengeometrie haben gute Ergebnisse gezeigt. Ein Vergleich mit einer vollaufgelösten Referenzsimulation zeigte Abweichungen der globalen Sherwood-Zahl für die Transferspezies von unter 4%.

Abstract

The high Schmidt-number problem is considered at single, free-rising gas bubbles with an additional chemical reaction of the transfer species. In such a case the concentration boundary layer is significantly smaller than the flow boundary layer. For Direct Numerical Simulation, this requires the use of very fine grids, which increase the simulation effort. To counteract this, previous works have used a subgrid-scale model for species transport, allowing to use coarser grids. The inaccurately predicted mass transport is corrected by intervention of the model. This is done by adjusting flux and source terms in the simulation. The subgrid-scale model was created by means of a data-driven approach using a neural network. The necessary training data was obtained from simulations of a simplified problem. Based on this approach, a data-driven subgrid-scale model is also used in this thesis. The process of training data generation is further simplified. A one-dimensional model is derived, with which the training data is calculated. This reduces the effort for data generation. Instead of a decay reaction for the transfer species, in this work the reaction of the transfer species with one species dissolved in the liquid is considered, thus two additional species occur. To facilitate the process of network training, a reduction of the training data set is performed to filter out redundant data. In this thesis, two reduction algorithms are developed, compared and finally combined. First tests of the subgrid-scale model on a simple bubble shape have shown good results. A comparison with a fully resolved reference simulation has shown deviations of the global Sherwood number for the transport species of less than 4%.



Masterarbeit von Herrn Alexander Kiefer
Matr.-Nr. 1687040

**Datenbasierte Subgridskalen-Modellierung reaktiver Konzentrationsgrenzschichten an
freiaufsteigenden Einzelblasen**

**Data-driven subgrid-scale modeling of reactive concentration boundary layers at single rising
bubbles**

Mit der stetigen Zunahme von Rechenressourcen ergeben sich neue Möglichkeiten für die Entwicklung physikalischer Modelle in Strömungssimulationen skalenübergreifender Prozesse. Am Institut für mathematische Modellierung und Analysis werden zu diesem Thema datenbasierte Ersatzmodelle für Konzentrationsgrenzschichten entwickelt, welche die Untersuchung reaktiver Blasenströmungen deutlich genauer und effizienter machen sollen. In der vorliegenden Arbeit geht es darum, vorhandene Modellierungsansätze für den physikalischen Stoffübergang auf eine Reaktion mit mehreren Komponenten zu erweitern. Konkret soll eine Reaktion untersucht werden, bei der die Übergangsspezies A mit einer Bulkspezies B zu einem gewünschten Produkt P reagiert, kurz $A + B \rightarrow P$. Das Programm zur Bearbeitung des Themas umfasst die folgenden Schwerpunkte:

1. Die Auswahl eines geeigneten Ersatzproblems und den damit einhergehenden globalen Parametern (z. B. Reynoldszahl, Seitenverhältnis, etc.) sowie das Festlegen der Parameterbereiche anhand der im *SPP 1740* entwickelten Reaktionssysteme.
2. Die Durchführung einer Parameterstudie mittels OpenFOAM. Dabei soll der zuvor festgelegte Parameterbereich möglichst effizient durch automatisierte Simulationen abgetastet werden.
3. Die Entwicklung eines Algorithmus zur gezielten Datenreduktion. Das Ziel des Algorithmus ist die Reduktion der Daten anhand zu erarbeitender Qualitätsmetriken, um die spätere Datenanalyse bzw. das Erstellen von Ersatzmodellen zu erleichtern.
4. Das Erstellen von datenbasierten Ersatzmodellen. Dazu gehört die Auswahl möglicher Modellierungsgrößen, z.B. motiviert durch Grenzschichttheorie, die Reduktion dieser Modellierungsgrößen mithilfe von Auswahlalgorithmen (z.B. *sequential backward selection*) und die datenbasierte Optimierung der Modellparameter.
5. Die Validierung der Ersatzmodelle auf Basis direkt-numerisch berechneter Grenzschichten an realistischen Blasenformen.

Mitbetreuer: Dipl. Ing. Andre Weiner, M. Sc. Dennis Hillenbrand

Profilbereich: Thermo-Fluids and Interfaces, Mathematische Modellierung und Analysis

Darmstadt, den 11. 9. 2019

Prof. Dr. Dieter Bothe

Darmstadt, den 9. 9. 19

Prof. Dr.-Ing. Peter Stephan

Inhaltsverzeichnis

1	Einleitung	5
2	Modellierung	8
2.1	Strömungsdynamik	8
2.2	Stofftransport	9
2.3	Diskretisierung	10
2.4	Arbeitsweise des Subgridskalen-Modells	11
2.5	Berechnung des Subgridskalen-Modells	15
3	Datengenerierung	17
3.1	Ersatzproblem	17
3.2	Numerisches Setup	17
3.3	Parameterbereich	19
3.4	Größe des Berechnungsgebiets	20
3.5	Räumliche Auflösung	23
3.6	1D-Modell	25
4	Datenreduktion	39
4.1	Übersicht	39
4.2	Round-Verfahren	39
4.3	Metrik-Verfahren	40
4.4	Vergleich	41
5	Datenbasierte Ersatzmodelle	45
5.1	Ablauf	45
5.2	Datenanalyse	46
5.2.1	Grundlagen	46
5.2.2	Feature-Auswahl	47
5.3	Netztraining	51
6	Validierung	55
7	Fazit	57
	Symbolverzeichnis	58
	Literaturverzeichnis	60
	Anhang	63

1 Einleitung

Stofftransport kommt in vielen Situationen vor, sei es bei alltäglichen Dingen wie Tee kochen oder bei industriellen Anwendungen wie Blasensäulenreaktoren [13]. Letztere werden in der chemischen Industrie benutzt, um verschiedene Stoffe miteinander in Kontakt zu bringen, sodass diese miteinander reagieren können. In einem Reaktor befindet sich eine Flüssigkeit, die entweder selbst einer der Reaktanten ist oder in der der Reaktant (im Folgenden als Spezies B bezeichnet) gelöst ist. Am Boden des Reaktors wird ein zweiter Reaktant (nachfolgend als Spezies A bezeichnet) eingebracht. Dieser ist gasförmig und steigt als ein Schwarm von Blasen in der Flüssigkeit auf. Aufgrund der Konzentrationsunterschiede zwischen Blase und Flüssigkeit für den gasförmigen Stoff diffundiert dieser an der Phasengrenze in die Flüssigkeit und reagiert mit Spezies B.

In Abbildung 1 ist beispielhaft ein Ausschnitt eines solchen Reaktors gezeigt. Wie zu erkennen ist, liegen mehrere Blasen unterschiedlicher Größe und Form vor. Diese können miteinander interagieren und sich beeinflussen. So kann der Nachlauf einer Blase Einfluss auf eine dahinter liegende Blase haben oder Blasen können kollidieren. Die Strömung und der Stofftransport sind dementsprechend komplex. Dies macht die numerische Strömungssimulation eines Blasensäulenreaktors, zum Beispiel um diesen in Hinsicht auf die Ausbeute an Reaktionsprodukt hin zu optimieren, sehr aufwändig.



Abbildung 1: Ein Blasensäulenreaktor (Bild 6 (b) in [2]).

Ein erster Schritt ist die Komplexität zu verringern, indem nur Einzelblasen betrachtet werden. So lassen sich grundlegende Kenntnisse über Zusammenhänge und Mechanismen für die Kopplung zwischen Fluidodynamik und Stoffübergang gewinnen. Die Numerik kann hierfür ein gutes Werkzeug sein, da experimentelle Methoden immer den Einschränkungen der Messtechnik und Messgenauigkeit unterliegen. So wurde z.B. in [11] eine oberflächenaktive Substanz verwendet um die Stoffverteilung in der Strömung sichtbar zu machen. Diese wiederum hat aber auch Einfluss auf den Stoffübergang und die Form und Geschwindigkeit der Blase.

Aber auch bei der Betrachtung von Einzelblasen ergeben sich Herausforderungen. So liegt ein zweiphasiges Problem vor. Auch die Strömung innerhalb der Blase kann Auswirkungen haben. Diese ebenfalls abzubilden erhöht den Simulationsaufwand, da auch das Innere der Blase und die Interaktion an der Phasengrenzfläche berechnet werden müssen. An dieser Stelle sei darauf hingewiesen, dass in dieser Arbeit die Strömung innerhalb der Blase vernachlässigt wird. Es wird also nur ein einphasiges Problem betrachtet und es wird nur die Umströmung der Blase simuliert. Die Form der Blase und deren Aufstiegs geschwindigkeit werden vorgegeben.

Die Fluidodynamik wird soweit vereinfacht, weil das Augenmerk dieser Arbeit auf einer anderen Problemstellung liegt. Aufgrund der Relativgeschwindigkeit zwischen der bewegten Blase und der ruhenden Flüssigkeit, sowie deren Viskosität, wird in einem engen Bereich um die Blase die Flüssigkeit verdrängt und sozusagen mitgerissen. Es kommt zu der Ausbildung einer dynamischen Grenzschicht mit Grenzschichtdicke δ_d um die Blase. In diesem Bereich hat die Flüssigkeit eine Geschwindigkeit $U_{fl} \neq 0$. Mit zunehmendem Abstand zur Blase wird dieser Einfluss kleiner und die Geschwindigkeit

nimmt ab. Um diesen Geschwindigkeitsgradienten richtig zu berechnen, muss die dynamische Grenzschicht mit genügend Zellen räumlich aufgelöst werden, was wiederum den Simulationsaufwand vergrößert. Aufgrund des Stoffübergangs an der Phasengrenze bildet sich noch eine zweite Grenzschicht. Diese beschreibt den Bereich der Flüssigkeit in dem Spezies A vorliegt. Die Grenzschichtdicke δ_c wird beeinflusst von der Diffusion, der Konvektion und von der Reaktion mit Spezies B. Analog dazu werden auch die Stoffverteilung von Spezies B und der Produktspezies P durch die Reaktion beeinflusst und es bilden sich Bereiche, in denen Gradienten für diese Spezies vorliegen. Die Speziesgrenzschicht muss ebenfalls aufgelöst werden, um die Stoffverteilung genau zu berechnen.

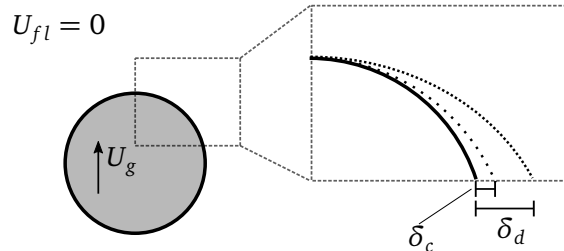


Abbildung 2: Schematische Darstellungen der dynamischen Grenzschichtdicke δ_d und der Stoffgrenzschichtdicke δ_c um eine mit der Geschwindigkeit U_g aufsteigenden Blase in ruhender Flüssigkeit f_l .

In welchem Größenverhältnis diese beiden Grenzschichten zueinander stehen, hängt von der Schmidt-Zahl Sc ab. Diese beschreibt das Verhältnis von diffusivem Impulstransport zu diffusivem Stofftransport und ist definiert als der Quotient von kinematischer Viskosität ν der Flüssigkeit und Diffusionskoeffizient D der jeweiligen Spezies in der Flüssigkeit, siehe Gleichung (1).

$$Sc = \frac{\nu}{D} \quad (1)$$

Bei großen Schmidt-Zahlen (z.B. $Sc = 500$) ist nun die Stoffgrenzschichtdicke δ_c viel kleiner als die dynamische Grenzschichtdicke δ_d , siehe Gleichung (2), entnommen aus [9].

$$\delta_c \propto \delta_d Sc^{-1/2} \quad (2)$$

Dieser Sachverhalt ist schematisch in Abbildung 2 dargestellt. Der Stofftransport findet auf einer viel kleineren räumlichen Skala statt und für eine genaue Berechnung dessen wird bei Direkter Numerischer Simulation eine sehr feine Auflösung des phasengrenznahen Bereichs benötigt. Dies vergrößert den Rechenaufwand für die Simulation.

Bisherige Lösungsansätze

In [1] wird der Stoffübergang einer gasförmigen Spezies in eine Flüssigkeit während des Gasblasenaufstiegs betrachtet, dabei treten keine chemischen Reaktionen auf. Es wird der Ansatz gewählt das Berechnungsgebiet in zwei Teile zu zerlegen. In einen Bereich der Flüssigkeit nahe der Gasblase und einen Bereich mit größerem Abstand zur Grenzfläche. In dem Bereich fern der Gasblase wird der Stofftransport auf einem für die Strömungsdynamik ausgelegtem Gitter berechnet. In dem Bereich nahe der Gasblase wird die Stoffverteilung normal zur Grenzfläche rein über ein aus analytischen Überlegungen hergeleitetem Modell berechnet. An der gemeinsamen Grenze beider Gebiete wird der Stofftransport über einen Quell- bzw. Senkenterm miteinander gekoppelt. Das analytische Modell ist aus Grenzschichtüberlegungen abgeleitet und beschreibt die eindimensionale Ausbreitung der Transferspezies.

Eine andere Art von Subgridskalen-Modell (englisch subgrid-scale-model, abgekürzt SGS-Modell) für den Stofftransport wird in [9] und [10] verwendet. Diese Art von Modell ist an die Finite-Volumen-Methode angepasst. Es muss nur die dynamische Grenzschicht aufgelöst werden, was die Simulationszeit verringert im Vergleich zu Simulationen, bei denen die Speziesgrenzschicht aufgelöst wird. Stofftransport und Strömungsdynamik werden auf dem für die Strömungsdynamik ausgelegtem Gitter berechnet, wobei die Stoffverteilung ungenau wird. Das SGS-Modell nutzt zellgemittelte Werte aus der Simulation, um Korrekturen an Flusstermen durchzuführen und so diese Ungenauigkeiten zu beheben. Wie dies geschieht ist in Kapitel 2.4 erläutert. Ein solches Subgridskalen-Modell kann auf analytischem Wege hergeleitet werden. Dieser Weg ist für einfache Blasengeometrien gangbar. Komplizierte Reaktionen der Spezies können ein Problem darstellen.

So wird in [9] ein SGS-Modell analytisch aus einer eindimensionalen Konvektions-Diffusions-Gleichung hergeleitet. Dieser Fall betrachtet den Stoffübergang einer einzigen Spezies ohne Auftreten einer Reaktion. Das Problem wird so weit vereinfacht, dass keine Zeitabhängigkeit vorhanden ist und der Stoffübergang an einer planaren Grenzfläche beschrieben wird.

Eine andere Möglichkeit für die Erstellung eines Subgridskalen-Modells ist ein datenbasierter Ansatz. Dieses Vorgehen erlaubt eine breitere Anwendung als speziell entwickelte analytische Subgridskalen-Modelle. Mit Trainingsdaten aus einfachen Geometrien wird per Machine Learning Algorithmen ein SGS-Modell berechnet. So wird in [10] für den Fall einer Einzelblase mit sphärischer oder ellipsoider Form dieser Ansatz gewählt. Außerdem liegt eine einfache Zerfallsreaktion, Reaktion erster Ordnung (vergleichbar mit radioaktiven Zerfallsprozessen), der Transferspezies A vor. Das SGS-Modell wird mit Hilfe eines sogenannten neuronalen Netzes berechnet. Dieses wird in das Simulationsprogram OpenFOAM eingebunden und nimmt Korrekturen bei der Lösung der Stofftransportgleichung vor.

Zielstellung

Ziel dieser Arbeit ist es, den zuvor genannten Ansatz eines datenbasierten SGS-Modells auf den Fall einer Reaktion zweiter Ordnung zu erweitern. Bei dieser Reaktion reagieren zwei Edukte (Spezies A und B) zu einem Produkt (Spezies P). Es gibt nunmehr drei Stofftransportgleichungen, die über den Reaktionsterm miteinander gekoppelt sind. Das SGS-Modell muss demnach ebenfalls auf drei Spezies erweitert werden. Die Berechnung des SGS-Modells mit neuronalen Netzen erfordert Trainingsdaten. Bei deren Generierung entstehen große Datenmengen. Ein weiteres Ziel dieser Arbeit ist die Entwicklung eines Algorithmus, um diese Datenmenge auf ein brauchbares Maß, mit möglichst geringem Informationsverlust, zu reduzieren.

Aufbau der Arbeit

Um dem Leser/der Leserin vorab einen roten Faden an die Hand zu geben, soll an dieser Stelle kurz der Aufbau dieser Ausarbeitung angesprochen werden. Das Problem unterschiedlicher Skalen bei großen Schmidt-Zahlen und die Zielsetzung wie dieses gelöst werden soll, wurden schon dargelegt.

In Kapitel 2 werden zuerst die grundlegenden Modellierungen zur numerischen Berechnung der Strömungsdynamik und des Stofftransports aufgeführt. Ausgehend von diesen wird ein detaillierter Blick auf die Arbeitsweise des Subgridskalen-Modells geworfen. Es wird gezeigt, an welcher Stelle bei einer unzureichenden Auflösung die Ungenauigkeiten entstehen. Darauf aufbauend wird gezeigt, wie das SGS-Modell diese Ungenauigkeiten korrigieren kann. Im letzten Abschnitt wird auf die Berechnungsgrundlage für das SGS-Modell eingegangen.

Das nächste Kapitel behandelt die Generierung von Trainingsdaten, um das SGS-Modell berechnen zu können. Es wird der Parameterbereich, den das spätere Modell abdecken soll, definiert. Die Trainingsdaten werden durch numerische Simulation eines vereinfachten Ersatzproblems berechnet. Dieses wird näher spezifiziert und die notwendigen Vorarbeiten für die numerische Simulation werden durchgeführt. Aufgrund der numerischen Ergebnisse wird das Ersatzmodell weiter vereinfacht auf einen eindimensionalen Ansatz für den Stofftransport. Dieser soll die Erzeugung von Trainingsdaten beschleunigen.

In Kapitel 4 wird das zweite Ziel dieser Arbeit, die Entwicklung eines Algorithmus für die Datenreduktion, angegangen. Es werden zwei verschiedene Verfahren beleuchtet, ihre Funktionsweise vorgestellt und ihre Eigenschaften miteinander verglichen.

Kapitel 5 beschäftigt sich mit der Berechnung der SGS-Modelle. Es werden zwei verschiedene Datensätze an Trainingsdaten verwendet, um damit vier Modell-Varianten zu berechnen. Die Datensätze werden mit den Methoden aus Kapitel 3 und 4 erzeugt und bearbeitet. In diesem Kapitel werden die Datensätze analysiert, um die Eingabegrößen für die SGS-Modelle zu bestimmen.

In Kapitel 6 werden die berechneten SGS-Modelle anhand eines Validierungsfalles auf ihre Genauigkeit hin untersucht und miteinander verglichen.

Im letzten Kapitel wird der Inhalt dieser Arbeit kurz zusammengefasst und ein Ausblick auf die nächsten Schritte gegeben.

2 Modellierung

Dieses Kapitel beschäftigt sich mit den für diese Arbeit nötigen Grundlagen. Es werden kurz die zugrunde liegenden Modelle für die Berechnungen erläutert. Ausgehend von den Grundgleichungen der Strömungsdynamik und des Stofftransports wird zur Diskretisierung letzterer übergegangen. Anschließend wird auf das Subgridskalen-Modell eingegangen und dessen Wirkungsweise bei der numerischen Simulation erklärt. Der letzte Abschnitt behandelt die Grundzüge der Berechnung des Subgridskalen-Modells mit Hilfe eines Machine Learning Algorithmus, also die Lösung des Regressionsproblems.

Konventionen

An dieser Stelle ein kleiner Vorgriff auf Notationskonventionen in diesem und den nachfolgenden Kapiteln, um an den entsprechenden Stellen den inhaltlichen Fluss nicht zu stören. In dieser Arbeit kommen sowohl dimensionsbehaftete (siehe Gleichungen (3)) als auch entdimensionalisierte Gleichungen, siehe Gleichung (4), vor. Dimensionslose Variablen werden mit einer Tilde gekennzeichnet, so zum Beispiel $\tilde{\beta}$. Dimensionsbehaftete Variablen haben diese nicht, β . Es treten auch mathematische Operatoren wie der Nabla-Operator ∇ und der Laplace-Operator Δ auf. Um die Schreibweise zu erleichtern, wird keine eigene Unterscheidung zwischen den dimensionsbehafteten und dimensionslosen Operatoren vorgenommen. Operatoren angewandt auf dimensionslose Variablen sind entdimensioniert, angewandt auf dimensionsbehaftete Größen sind sie dimensionsbehaftet. Die Operatoren werden analog mit den zugehörigen charakteristischen Werten entdimensionalisiert. Weiterhin kommen Abkürzungen vor. So bedeutet ∂_x ausgeschrieben $\frac{\partial}{\partial x}$. Für diese Abkürzung gilt die gleiche Konvention bezüglich der Dimensionalität wie für die zuvor genannten Operatoren. Außerdem bedeutet $\partial_x \beta|_l$ die Auswertung von $\partial_x \beta$ an der Stelle $x = l$. Fett geschriebene Variablen, \mathbf{g} , bezeichnen Vektoren. Variablen mit einem Balken, \bar{c}_A , bezeichnen gemittelte Werte.

2.1 Strömungsdynamik

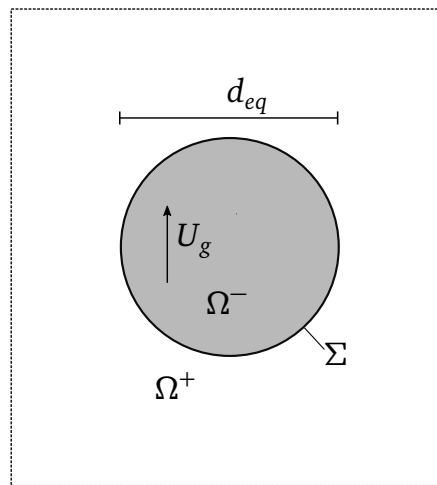


Abbildung 3: Skizze charakteristische Größen.

In Abbildung 3 ist eine aufsteigende Blase des Durchmessers d_{eq} und der Geschwindigkeit U_g skizziert. Das Innere der Blase wird mit Ω^- bezeichnet und ist die gasförmige Phase angereichert mit Spezies A. Das flüssige Medium, in dem die Blase aufsteigt, umfasst den mit Ω^+ gekennzeichneten Bereich. Die Phasengrenze wird durch die Grenzfläche Σ dargestellt. Als Vereinfachung werden in dieser Arbeit die Vorgänge im Inneren der Blase vernachlässigt und nur die Flüssigphase des Problems und die dazugehörige Seite der Phasengrenze Σ betrachtet. Der Einfachheit halber wird Ω^+ im Weiteren nur noch als Ω bezeichnet. Es handelt sich nun nur noch um ein einphasiges Problem und dementsprechend wird an dieser Stelle nicht auf Sprungbedingungen zwischen den Phasen eingegangen. Außerdem muss nur noch die Strömung des Mediums in Ω berechnet werden.

Grundlage der numerischen Strömungssimulation ist die Diskretisierung der Navier-Stokes-Gleichungen, welche die Strömungsdynamik beschreiben. Sie sind abgeleitet aus der Bilanzierung von Masse und Impuls und sind um das rheologische Modell der Newtonschen Fluide ergänzt. Gleichung (3) gibt diese für den inkompressiblen, dreidimensionalen Fall wieder. Die zu berechnenden Strömungsgrößen sind hierbei der Geschwindigkeitsvektor \mathbf{u} und der Druck p . Die Dichte ρ

und die dynamische Viskosität μ sind Stoffgrößen des Mediums. Mit \mathbf{g} wird die Gravitationsfeldstärke in vektorieller Form bezeichnet. $\frac{D}{Dt}$ ist eine totale Ableitung.

$$\begin{aligned}\nabla \cdot \mathbf{u} &= 0 \\ \rho \frac{D\mathbf{u}}{Dt} &= \rho \mathbf{g} - \nabla p + \mu \Delta \mathbf{u}\end{aligned}\quad (3)$$

Die Größen dieser Gleichungen sind dimensionsbehaftet. Durch Verwendung charakteristischer Werte für den jeweiligen Anwendungsfall lassen sich diese entdimensionalisieren. Dazu wird eine charakteristische Geschwindigkeit \hat{U} , eine charakteristische Länge \hat{L} und vorgreifend auf Kapitel 2.2 eine charakteristische Konzentration \hat{c} eingeführt. Die Entdimensionalisierungsvorschriften sind in Tabelle 1 aufgelistet. Die Festlegung der charakteristischen Größen erfolgt in Kapitel 3.2.

Zeit	Länge	Geschwindigkeit	Konzentration	Druck	Gravitationsfeldstärke
$\tilde{t} = \frac{t}{\hat{L}}$	$\tilde{x}_j = \frac{x_j}{\hat{L}}$	$\tilde{u}_j = \frac{u_j}{\hat{U}}$	$\tilde{c}_i = \frac{c_i}{\hat{c}}$	$\tilde{p} = \frac{p}{\rho \hat{U}^2}$	$\tilde{g}_j = \frac{g_j \hat{L}}{\hat{U}^2}$

Tabelle 1: Entdimensionalisierungsvorschrift.

Die Entdimensionalisierung erleichtert die Vergleichbarkeit mit anderen Fällen und reduziert die Anzahl möglicher Konfigurationen, da mehrere physikalische Größen in dimensionslosen Kennzahlen, wie der Reynoldszahl Re , zusammengefasst werden (siehe Tabelle 2). Unterschiedliche (nicht ähnliche) Fälle drücken sich in unterschiedlichen Kennzahlen aus und dementsprechend müssen nur diese variiert werden und nicht die physikalischen Größen. Die dimensionslosen Navier-Stokes-Gleichungen sind in Gleichung (4) dargestellt.

$$\begin{aligned}\nabla \cdot \tilde{\mathbf{u}} &= 0 \\ \frac{D\tilde{\mathbf{u}}}{D\tilde{t}} &= \tilde{\mathbf{g}} - \nabla \tilde{p} + \frac{1}{Re} \Delta \tilde{\mathbf{u}}\end{aligned}\quad (4)$$

Für die eindeutige Lösung dieser Gleichungen müssen nun noch Bedingungen für die zu berechnenden Größen Geschwindigkeit und Druck an allen Rändern des Berechnungsgebietes getroffen werden. Diese werden Randbedingungen genannt und sind problemspezifisch. Deswegen wird an dieser Stelle auf Kapitel 3.2 verwiesen. Da es sich um ein transientes Problem handelt, müssen Startzustände für das Geschwindigkeits- und Druckfeld definiert werden. Diese sogenannten Anfangsbedingungen werden ebenfalls in Kapitel 3.2 angegeben.

2.2 Stofftransport

Ziel der Simulation ist es, den Übergang eines Stoffes (Spezies A), der als Gasblase in einer Flüssigkeit aufsteigt, zu berechnen. In dem Fluid ist ein Stoff (Spezies B) gelöst, welcher unter einer vereinfacht dargestellten Reaktion ($A + B \rightarrow P$) zur Spezies P reagiert. Der Transport jeder Spezies wird durch eine eigene Transportgleichung für die jeweilige Konzentration beschrieben. Die Konzentration c_i eines Stoffes i ist die Stoffmenge N_i in $[mol]$ in einem bestimmten Volumen V . Die drei Spezies können sich durch die Transportmechanismen Diffusion und Konvektion in dem Berechnungsgebiet verteilen. Die Konzentrationsänderung wird außerdem durch die Reaktion der Stoffe miteinander beeinflusst und durch den Reaktionsterm R_i in der Gleichung (5) dargestellt. Dieser dient als „Quelle“ (Spezies P) bzw. als „Senke“ (Spezies A und B) für die jeweiligen Stoffe.

$$\frac{\partial c_i}{\partial t} = D_i \Delta c_i - \mathbf{u} \cdot \nabla c_i + R_i \quad (5)$$

Die vereinfachte Reaktion der Stoffe A und B ist eine Reaktion zweiter Ordnung. Der Reaktionsterm ist dementsprechend abhängig sowohl von Konzentration c_A als auch von c_B und negativ, da die Stoffe A und B verbraucht werden. Der Reaktionsterm ist für Spezies A und B identisch und in Gleichung (6) dargestellt.

$$R_A = R_B = -k c_A c_B \quad (6)$$

Im gleichen Maße wie Spezies A und B verbraucht werden, wird Spezies P bei der Reaktion gebildet. Die Berechnungsvorschrift für Reaktionsterm R_p ist dementsprechend bis auf das Vorzeichen gleich, siehe Gleichung (7).

$$R_p = k c_A c_B \quad (7)$$

Der Faktor k in den Gleichungen (6) und (7) ist der Geschwindigkeitskoeffizient der Reaktion und ein Maß für die Stärke der Reaktion. Die Diffusionskoeffizienten werden der Einfachheit halber für alle drei Spezies als gleich angenommen ($D_A = D_B = D_p = D$). Analog zur Navier-Stokes-Gleichung kann auch Gleichung (5) entdimensionalisiert werden, somit ergibt sich Gleichung (8). Bei dieser Entdimensionalisierung kommt nun auch die schon erwähnte charakteristische Konzentration zur Anwendung.

$$\frac{\partial \tilde{c}_i}{\partial \tilde{t}} = \frac{1}{Pe} \Delta \tilde{c}_i - \tilde{\mathbf{u}} \cdot \nabla \tilde{c}_i \pm Da \tilde{c}_A \tilde{c}_B \quad (8)$$

Durch die Entdimensionalisierung der Strömungs- und Stofftransportgleichungen treten in den entstehenden Gleichungen dimensionslose Kennzahlen auf. Diese sind in Tabelle 2 zusammengefasst. So tritt in den Impulsgleichungen, Gleichung (4), die Reynolds-Zahl Re auf. Diese beschreibt das Verhältnis von Trägheits- zu Zähigkeitskräften. Bei größer werdenden Re nimmt der Einfluss der Viskosität ab, was kleinere dynamische Grenzschichten bedeutet. In der Stofftransportgleichung (8) taucht die Péclet-Zahl Pe auf. Sie beschreibt das Verhältnis von Stofftransport per Konvektion zu Stofftransport per Diffusion. So überwiegt bei großen Pe die Konvektion und bei kleinen Pe die Diffusion. Die Re und die Pe sind über die Schmidt-Zahl Sc miteinander gekoppelt, siehe Gleichung (9). Wie schon in der Einleitung erläutert, ist bei großen Sc die Strömungsgrenzschicht dicker als die Stoffgrenzschicht. Außerdem tritt die Damköhler-Zahl Da auf. Sie ist eine dimensionslose Kennzahl, welche die Strömungszeitskala mit der Reaktionszeitskala ins Verhältnis setzt. Sie tritt in unterschiedlichen Definitionen auf. Bei einer Reaktion 2. Ordnung ergibt sich die in Tabelle 2 aufgeführte Variante.

$$Pe = Re Sc \quad (9)$$

Für eine eindeutige Lösung müssen auch für den Stofftransport für alle Spezies Randwerte vorgegeben werden. Ebenso sind die Anfangsverteilungen der drei Spezies zu definieren. Hier wird wieder auf Kapitel 3.2 verwiesen.

Reynolds-Zahl	Péclet-Zahl	Schmidt-Zahl	Damköhler-Zahl
$Re = \frac{\tilde{U} \tilde{L}}{\nu}$	$Pe = \frac{\tilde{U} \tilde{L}}{D}$	$Sc = \frac{\nu}{D}$	$Da = \frac{\tilde{L} k c}{\tilde{U}}$

Tabelle 2: Dimensionslose Kennzahlen.

2.3 Diskretisierung

Die Navier-Stokes-Gleichungen und die Stofftransportgleichungen gelten in jedem Punkt des Berechnungsgebietes. Die numerische Strömungssimulation geschieht mit der Finite-Volumen-Methode. Das Rechenggebiet wird in viele kleine Volumina, Kontrollvolumina KV, zerlegt. Für jedes dieser KV werden die Bilanzgleichungen (3 und 5) in Integralform verwendet. Der Fokus soll hier auf der Stofftransportgleichung (5) liegen. In [8] ist die Diskretisierung für die Navier-Stokes-Gleichungen beschrieben und wird hier analog für die Stofftransportgleichung behandelt. Mit Hilfe des Gaußschen Integralsatzes kann die integrale Form der Bilanzgleichung über ein KV umgeformt werden. Die entstehenden Oberflächenintegrale können nun auch als Summe der Integrale über die einzelnen Teilflächen f eines KV betrachtet werden, siehe Gleichung (10).

$$\sum_f \int_{S_f} \left(u_j c_i - D \frac{\partial c_i}{\partial x_j} \right) n_{fj} dS_f = \int_V \left(\frac{\partial c_i}{\partial t} \mp R_i \right) dV \quad (10)$$

Die Oberflächenintegrale können nun aufgeteilt werden in die diffusiven F_f^D und advektiven Flüsse F_f^A für die Seite f , wobei n_{fj} die j -te Komponente des Normalenvektors der Seite f ist.

$$\begin{aligned}
F_f^D &= \int_{S_f} \left(-D \frac{\partial c_i}{\partial x_j} \right) n_{fj} dS_f \\
F_f^A &= \int_{S_f} (u_j c_i) n_{fj} dS_f
\end{aligned} \tag{11}$$

Als Nächstes müssen diese Oberflächenintegrale approximiert werden. Aus der Mittelpunktregel und der Konvention, dass \mathbf{S}_f der Normalenvektor auf Seite f mit Betrag gleich dem Flächeninhalt S_f der Seite ist, folgt Gleichung (12).

$$\begin{aligned}
F_f^D &\approx -D (\nabla c_i)_f \cdot \mathbf{S}_f \\
F_f^A &\approx c_{i,f} \mathbf{u} \cdot \mathbf{S}_f
\end{aligned} \tag{12}$$

Analog wird das Volumenintegral approximiert. Hier wird der Wert des Integrals angenähert durch das Produkt des Mittelwerts des Integranten (über dem betrachteten KV und definiert am Mittelpunkt ξ der Zelle) mit dem Volumen V des KV. Es ergibt sich somit der Zusammenhang in Gleichung (13) für die rechte Seite.

$$\int_V \frac{\partial c_i}{\partial t} \mp R_i dV \approx \left[\left(\frac{\partial c_i}{\partial t} \right)_\xi \mp (R_i)_\xi \right] V \tag{13}$$

Damit lässt sich Gleichung (10) in Gleichung (14) überführen, dabei wird noch die Vereinfachung $\phi_f = \mathbf{u} \cdot \mathbf{S}_f$ genutzt.

$$\sum_f c_{i,f} \phi_f - D (\nabla c_i)_f \cdot \mathbf{S}_f \approx \left(\frac{\partial c_i}{\partial t} \right)_\xi V \mp (R_i)_\xi V \tag{14}$$

Die in der Gleichung (14) vorkommenden Gradienten werden schließlich mit finiten Differenzen approximiert. So kann z.B. die Ableitung eines Wertes Ψ am Ort der Fläche f zwischen zwei Zellen durch einen zentralen Differenzenquotienten zwischen den Mittelpunkten ξ_1 und ξ_2 der beiden Zellen dargestellt werden, siehe Gleichung (15).

$$\left(\frac{\partial \Psi}{\partial x} \right)_f \approx \frac{\Psi_{\xi_1} - \Psi_{\xi_2}}{x_{\xi_1} - x_{\xi_2}} \tag{15}$$

Hier zeigt sich der Konflikt zwischen Genauigkeit und Rechenaufwand. Durch weniger Zellen sinkt der Rechenaufwand, da für jede Zelle Gleichung (14) gelöst werden muss, die miteinander gekoppelt sind, aufgrund der Approximation der Gradienten. Jedoch wird, vor allem in Gebieten mit großen Gradienten, die Approximation dieser ungenauer. An diesem Konflikt setzt die Idee eines Subgridskalen-Modells an.

2.4 Arbeitsweise des Subgridskalen-Modells

In diesem Kapitel geht es um die Arbeitsweise des Subgridskalen-Modells. Dieses kann für einfache Beispiele analytisch hergeleitet werden (siehe z.B. [9]) oder, wie für diese Arbeit, auf Basis von *Machine Learning* Algorithmen erstellt werden.

Das Subgridskalen-Modell ermöglicht es eine gröbere räumliche Diskretisierung in der Nähe der Grenzfläche, also weniger Zellen, zu verwenden und damit die Simulation zu beschleunigen. Die dadurch entstehenden Fehler bei der Berechnung werden durch Korrekturterme behoben, welche das Subgridskalen-Modell liefert. Konkret auf diese Arbeit bezogen bedeutet dies, dass ein Gitter gewählt wird, welches eine ausreichende Auflösung für die Berechnung des Strömungsfeldes aufweist. Für die Berechnung der Stoffverteilung ist diese Auflösung zu gering ($\delta_c < \delta_d$), wodurch Fehler entstehen. Diese werden durch das Subgridskalen-Modell korrigiert.

Dieses Kapitel ist in zwei Teile geteilt. Der erste Abschnitt behandelt die Effekte der unzureichenden räumlichen Auflösung und veranschaulicht diese. Der zweite Teil beschäftigt sich mit der Frage, wie die auftretenden Fehler in einer mit

dem Simulationsprogramm, in diesem Fall OpenFOAM, verträglichen Weise korrigiert werden können.

Effekt unzureichender Auflösung

Dieser Abschnitt gibt die Erkenntnisse aus Kapitel 5.1 in [12] wieder. Von dort sind auch die hier verwendeten Abbildungen. Die Auswirkungen einer unzureichenden Auflösung auf die Berechnung der Konzentrationsverteilung wird dort anhand einer ähnlichen Problemstellung geschildert. Es wird eine eindimensionale Reaktions-Diffusions-Gleichung und die Reaktion $2 A + B \rightarrow P$ verwendet. In Gleichung (16) ist die Differentialgleichung für Spezies A dargestellt. Im Gegensatz dazu wird in dieser Arbeit eine Advektions-Diffusions-Gleichung mit Reaktion $A + B \rightarrow P$ verwendet, vergleiche Gleichung (5). Obwohl dies nicht die exakt gleichen Gleichungen sind, wird davon ausgegangen, dass sie ähnlich genug sind, um vergleichbar zu sein und die Erkenntnisse übertragen zu können.

$$\frac{d^2 c_A}{dx^2} = -k c_A^2 c_B \quad (16)$$

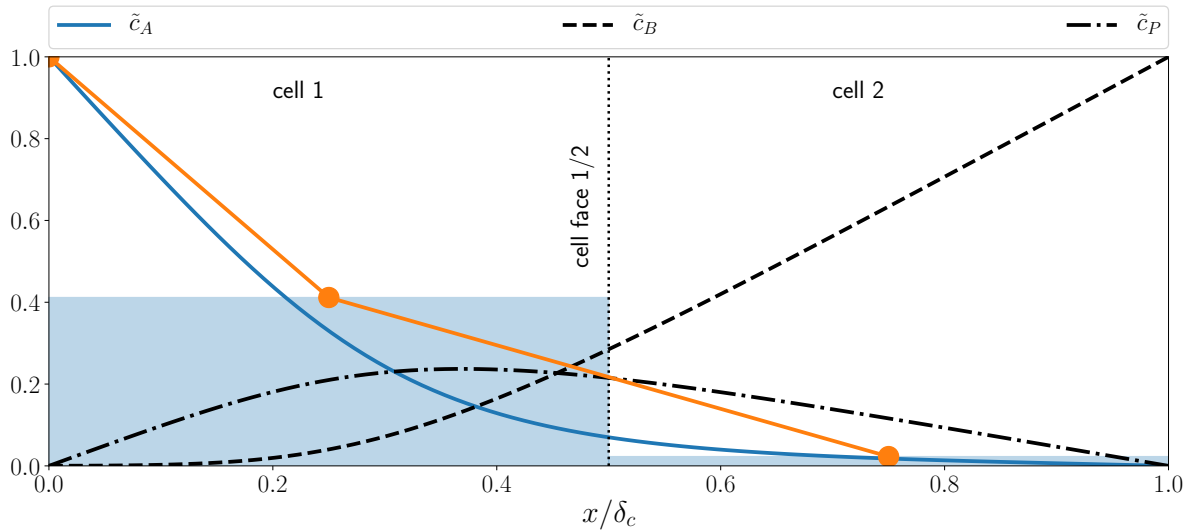


Abbildung 4: Effekt unzureichender Auflösung auf Spezies A (aus [12]).

In diesem Abschnitt wird verdeutlicht, wie sich eine unzureichende räumliche Auflösung auswirkt. So ist in Abbildung 4 der Verlauf der Konzentration der Spezies A normal zur Grenzfläche als blaue Linie dargestellt. Wie zu erkennen ist, liegt in diesem Beispiel die Konzentrationsgrenzschicht komplett innerhalb der ersten zwei Zellen des Simulationsgitters. Die numerische Lösung auf diesem Gitter ergibt zellzentrierte Werte \tilde{c}_A^{num} für Spezies A, welche über die Zelle konstant sind (hier als orangefarbene Punkte dargestellt). Um die diffusiven und advektiven Flüsse für Zelle 1 berechnen zu können, wird mit Hilfe dieser zellzentrierten Werte $\tilde{c}_A^{num}|_{x/\delta_c=0.25}$ und $\tilde{c}_A^{num}|_{x/\delta_c=0.75}$ die Konzentration auf der Zellfläche bei $x/\delta_c = 0.5$ linear interpoliert (orangefarbene Linie). Es ist deutlich zu erkennen, dass $\tilde{c}_A^{num}|_{x/\delta_c=0.5}$ größer ist als $\tilde{c}_A|_{x/\delta_c=0.5}$ und somit nach Gleichung (12) der advective Fluss betragsmäßig überschätzt wird. Auch ist zu erkennen, dass die numerisch berechneten Konzentrationsableitungen $d_x \tilde{c}_A^{num}$ bei $x/\delta_c = 0$ und $x/\delta_c = 0.5$ nicht mit $d_x \tilde{c}_A$ an den gleichen Stellen übereinstimmen. Der diffusive Fluss wird demnach bei $x/\delta_c = 0$ unter- und bei $x/\delta_c = 0.5$ überschätzt (vergleiche Gleichung (12)).

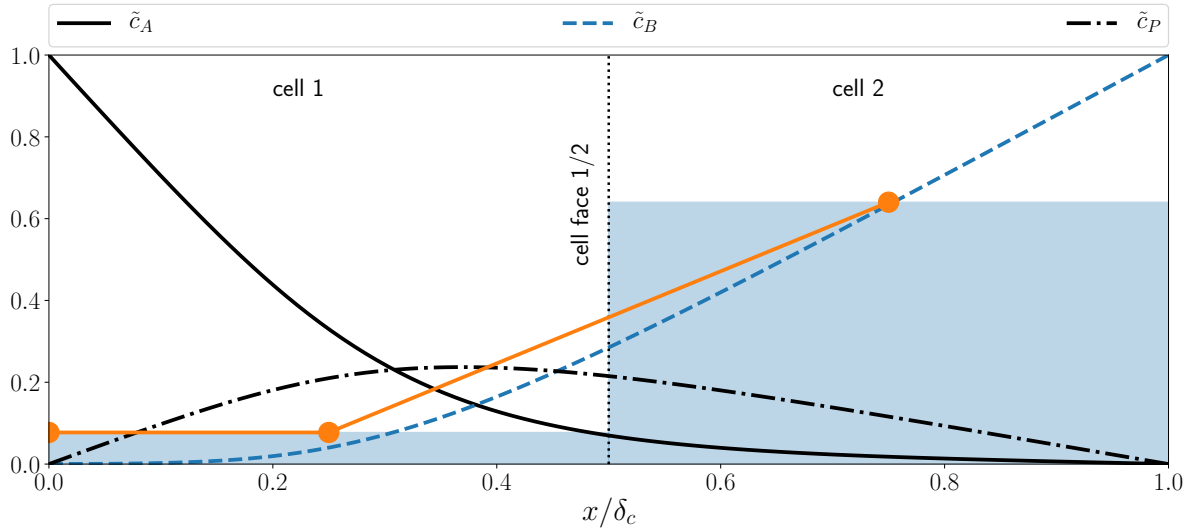


Abbildung 5: Effekt unzureichender Auflösung auf Spezies B (aus [12]).

In Abbildung 5 ist die gleiche Problemstellung dargestellt, jedoch liegt der Fokus nun auf der Spezies B. Der Verlauf von \tilde{c}_B ist als blaue Linie hervorgehoben. Die numerisch berechneten Werte in den Zellmittelpunkten ($\tilde{c}_B^{num}|_{x/\delta_c=0.25}$ und $\tilde{c}_B^{num}|_{x/\delta_c=0.75}$) sind analog als orangene Punkte dargestellt. Im Gegensatz zu Spezies A müssen bei Spezies B nur die Flüsse über die Fläche zwischen Zelle 1 und 2 betrachtet werden. Aufgrund der vorgegebenen Randbedingung an der Grenzfläche Σ (bei $x/\delta_c = 0$) verschwindet der Gradient dort. Die Konzentration \tilde{c}_B^{num} ist zwar größer als \tilde{c}_B an der Stelle $x/\delta_c = 0$, aber an dieser Stelle verschwindet die Geschwindigkeit normal zur Grenzfläche und damit auch der advective Fluss. Für die Flüsse an der Fläche zwischen den beiden Zellen werden analog zu Spezies A auch für Spezies B die Konzentration \tilde{c}_B^{num} und die Ableitung $d_x \tilde{c}_B^{num}$ aus den zellzentrierten Werten berechnet. Es ist zu erkennen, dass der advective Fluss über diese Fläche überschätzt und der diffusive Fluss unterschätzt wird.

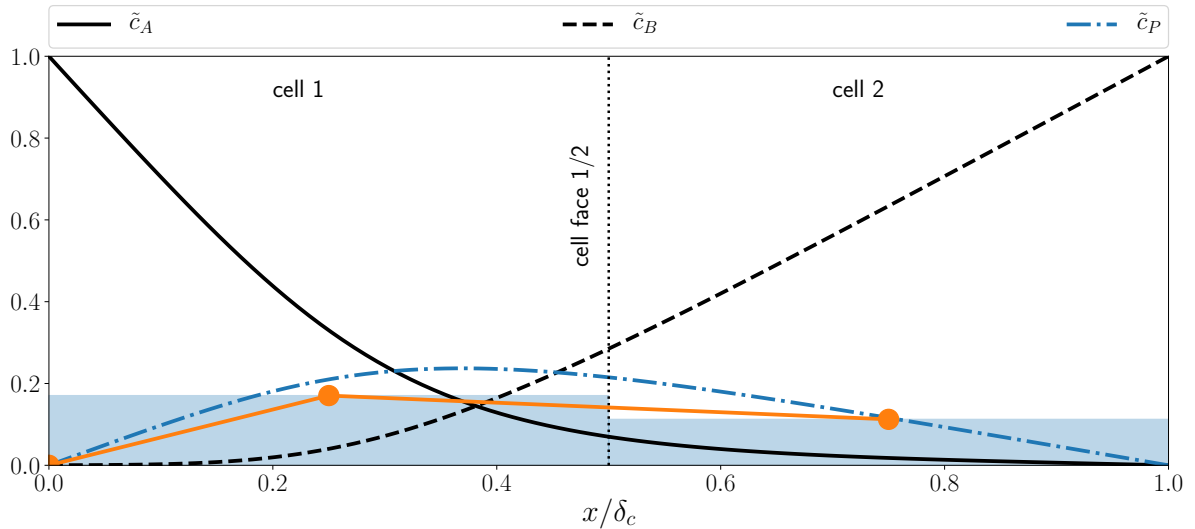


Abbildung 6: Effekt unzureichender Auflösung auf Spezies P (aus [12]).

In Abbildung 6 steht die Spezies P im Fokus. Deren Konzentrationsverlauf ist mit der blauen Linie dargestellt. Für Spezies P ergibt sich im Gegensatz zu den anderen beiden Spezies ein nicht monotoner Verlauf. Für Spezies P müssen ebenso wie für Spezies B nur die Flüsse über die Fläche zwischen Zelle 1 und 2 betrachtet werden. Analog zu den Betrachtungen für Spezies B ergeben sich auch hier Abweichungen zwischen den numerisch berechneten Werten und dem Konzentrationsverlauf und dessen Ableitung. Der diffusive und advective Fluss weichen dementsprechend ebenso ab.

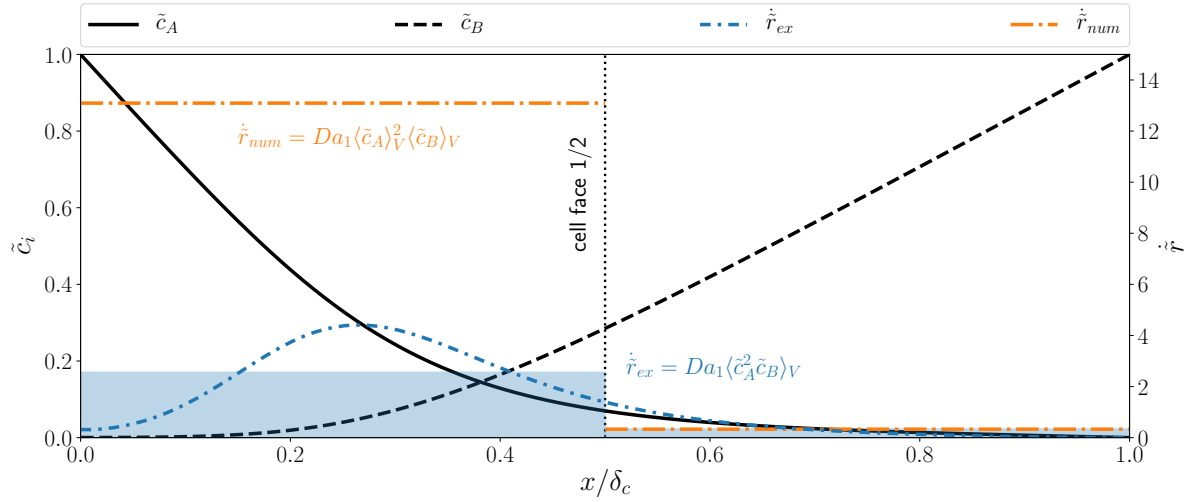


Abbildung 7: Effekt unzureichender Auflösung auf Reaktionsterm (aus [12]).

Neben den drei Spezies muss auch der Reaktionsterm betrachtet werden. In diesen gehen die berechneten Werte der Konzentrationen \tilde{c}_A^{num} und \tilde{c}_B^{num} ein. Demnach wird auch dieser als über die Zelle konstanter Mittelwert behandelt. In Abbildung 7 ist der Verlauf des Wertes des Reaktionsterms als Produkt der Konzentrationsverläufe der Spezies A und B und der Damköhler-Zahl als blaue Linie dargestellt. Der Mittelwert über die Zelle dieses Produktes ist als die blau schattierte Fläche gekennzeichnet. Es ist zu erkennen, dass der numerisch berechnete Wert (orange) sich deutlich davon unterscheidet. Allgemein ausgedrückt muss das Produkt der Mittelwerte über eine Zelle nicht gleich dem Mittelwert der Produkte sein, was zu Ungenauigkeiten bei der Berechnung des Reaktionsterms führen kann.

Dieser Abschnitt verdeutlicht, dass das Subgridskalen-Modell für die drei Spezies A, B und P sowohl die diffusiven und advektiven Flüsse als auch den Reaktionsterm korrigieren muss, um genaue Stoffverteilungen zu berechnen.

Korrektur in OpenFOAM

Dieser Abschnitt bezieht sich auf [6], dort sind auch die Herleitungen für die Korrekturterme nachlesbar (siehe dort Anhang B). Auch in [12] wird auf diesen Aspekt eingegangen.

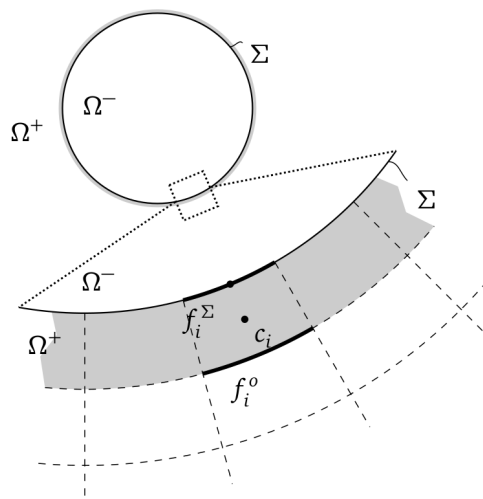


Abbildung 8: Skizze der ersten Zellen an der Grenzfläche (aus [12]).

Wie in dem vorherigen Abschnitt erläutert, tritt bei unzureichender räumlicher Auflösung eine Unter- bzw. Überschätzung der diffusiven/advektiven Flüsse auf. Auch die Berechnung des Reaktionsterms kann ungenau sein. Durch Verwendung des SGS-Modells soll dieser Genauigkeitsverlust verringert werden. Dies geschieht durch die Korrektur der diffusiven

und advektiven Flüsse und des Reaktionsterms in der OpenFOAM Simulation. Die Korrektur wird nicht im gesamten Berechnungsgebiet vorgenommen, sondern nur an der Zellfläche f_i^Σ , die numerisch die Grenzfläche Σ darstellt, und der ersten gegenüberliegenden Fläche f_i^o (siehe Abbildung 8). Es wird gefordert, dass die erste Zellreihe, wie in der Abbildung ersichtlich, an der Grenzfläche Σ ausgerichtet ist.

Zur Erinnerung wird an dieser Stelle auf die in Abschnitt 2.3 dargestellten Gleichungen für den diffusiven und advektiven Fluss hingewiesen, Gleichung (12). Für die Korrektur des diffusiven Flusses wird nicht direkt der fehlerhafte Konzentrationsgradient korrigiert, sondern es wird der Diffusionskoeffizient modifiziert. Durch diesen modifizierten Koeffizienten $D_{f_i^*}$ wird der Fluss auf den korrekten Wert skaliert. Dieser Koeffizient setzt sich aus dem vorgegebenen Diffusionskoeffizienten D (Stoffkennwert) und einem Skalierungsfaktor zusammen. Der Faktor setzt die, vom SGS-Modell berechnete, Normalableitung der Konzentration ins Verhältnis zur von OpenFOAM (*num*) berechneten Normalableitung der Konzentration, siehe Gleichung (17). Hierbei ist zu beachten, dass nur für die Übergangsspezies A die Diffusion an der Grenzfläche korrigiert werden muss. Für Spezies B und P ist als Randbedingung der numerischen Berechnung an dieser Stelle die Normalableitung zu Null vorgegeben.

$$D_{f_i^*} = D \frac{(\partial_n c)_{f_i^*}^{SGS}}{(\partial_n c)_{f_i^*}^{num}} \quad (17)$$

Die Korrektur des advektiven Flusses erfolgt analog zu der des diffusiven Flusses. Korrigiert wird in diesem Fall der Flächenfluss $\phi = \mathbf{S} \cdot \mathbf{u}$ auf der jeweiligen Fläche. Für die Korrektur des Advektionsterm muss nur die erste Zellfläche gegenüber der Grenzfläche f_i^o betrachtet werden, da die Geschwindigkeit normal zur Grenzfläche verschwindet und demnach gilt $\phi_{f_i^\Sigma} = \mathbf{S}_{f_i^\Sigma} \cdot \mathbf{u}_{f_i^\Sigma} = 0$.

$$\phi_{f_i^{\Sigma,o}} = \phi_{f_i^{\Sigma,o}}^{num} \frac{c_{f_i^{\Sigma,o}}^{SGS}}{c_{f_i^{\Sigma,o}}^{num}} \quad (18)$$

Wie schon in Abbildung 7 gezeigt, muss auch der Reaktionsterm korrigiert werden. Das Produkt der Mittelwerte der Konzentrationen der Spezies A und B unterscheidet sich von dem Mittelwert des Produkts der Konzentrationen. Für die Korrektur des Reaktionsterm kann direkt der fehlerhafte Term durch den vom Subgridskalen-Modell berechneten Wert ersetzt werden.

Zusammenfassend werden für die Korrekturen folgende Größen benötigt. Für Spezies A muss die Normalableitung der Konzentration an der Grenzfläche $\partial_{\tilde{n}} \tilde{c}_A|_\Sigma$ und bei einem bestimmten Abstand \tilde{l} berechnet werden, $\partial_{\tilde{n}} \tilde{c}_A|_{\tilde{l}}$. Außerdem wird die Konzentration bei einem Abstand \tilde{l} gebraucht, $\tilde{c}_A|_{\tilde{l}}$. Für Spezies B und P werden, bis auf die Normalableitung an der Grenzfläche, die gleichen Größen für die jeweiligen Konzentrationen benötigt. Für den Reaktionsterm wird der Mittelwert des Produkts der Konzentration von Spezies A und B bis zu einem bestimmten Abstand \tilde{l} benötigt, $\overline{\tilde{c}_A \tilde{c}_B}|_{\tilde{l}}$. Diese Größen muss das Subgridskalen-Modell bereitstellen, um eine Korrektur zu ermöglichen. In Tabelle 3 sind diese zusammengefasst.

$$\partial_{\tilde{n}} \tilde{c}_A|_\Sigma \quad \partial_{\tilde{n}} \tilde{c}_A|_{\tilde{l}} \quad \tilde{c}_A|_{\tilde{l}} \quad \partial_{\tilde{n}} \tilde{c}_B|_{\tilde{l}} \quad \tilde{c}_B|_{\tilde{l}} \quad \partial_{\tilde{n}} \tilde{c}_P|_{\tilde{l}} \quad \tilde{c}_P|_{\tilde{l}} \quad \overline{\tilde{c}_A \tilde{c}_B}|_{\tilde{l}}$$

Tabelle 3: Für die Korrektur benötigte Größen bei Benutzung des SGS-Modells.

2.5 Berechnung des Subgridskalen-Modells

Das Vorgehen, das Subgridskalen-Modell durch einen datenbasierten Ansatz (*Machine Learning*) zu realisieren, führt auf ein Regressionsproblem hin. Dabei muss durch Regression ein Zusammenhang zwischen den Eingabedaten, z.B. dem Abstand der betrachteten Zellfläche zur Grenzfläche, und den für die Korrektur der Simulation benötigten Größen, siehe Tabelle 3, hergestellt werden. In dieser Arbeit wird für die Lösung des Regressionsproblems ein sogenanntes neuronales Netz verwendet. Im Bereich des *Machine Learnings* werden die Eingabedaten als Features und die Ausgabedaten als Label bezeichnet.

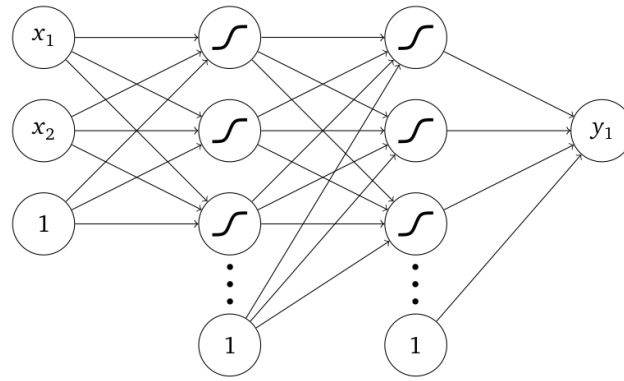


Abbildung 9: Beispielhafte Skizze eines neuronalen Netzes übernommen aus [12].

In Abbildung 9 ist beispielhaft ein neuronales Netz skizziert. Ein neuronales Netz besteht aus mehreren Schichten, die sich aus einzelnen Neuronen zusammensetzen. Die Anzahl der Neuronen in der Eingabeschicht wird durch die Anzahl der Features bestimmt. Dazu kommt noch ein Neuron als *bias unit*, welches als *offset* dient. In dieser Arbeit wird für jedes Label ein eigenes neuronales Netz trainiert. Dies bedeutet, dass die Ausgabeschicht nur aus einem einzigen Neuron besteht. Zwischen Eingabe- und Ausgabeschicht kann es mehrere versteckte Schichten geben. Die Anzahl an Neuronen pro versteckter Schicht ist frei wählbar, genauso wie die Anzahl an versteckten Schichten. Sind zu wenige Neuronen in den versteckten Schichten, so kann das neuronale Netz die Abhängigkeit zwischen den Features und dem Label nicht gut abbilden (*underfitting*). Sind jedoch zu viele Neuronen vorhanden passt sich das Netz auch an kleine Störungen (z.B. Rauschen) in den Daten an (*overfitting*). Die Anzahl an versteckten Schichten und Neuronen pro Schicht kann mit einer Hyperparameteroptimierung bestimmt werden. In sogenannten *feed forward*-Netzen sind die Neuronen nur mit Neuronen einer darauffolgenden Schicht verbunden. Die Verbindungen der Neuronen sind gewichtet. Diese Gewichte bilden die freien Parameter eines neuronalen Netzes, die es ermöglichen den Zusammenhang zwischen den Features und dem Label zu approximieren. Außerdem hat jedes Neuron eine sogenannte Aktivierungsfunktion, wie z.B. die Sigmoid-Funktion. Diese Aktivierungsfunktion sorgt dafür, dass das Neuron nur bei bestimmten Eingaben “aktiv” wird bzw. steuert in welcher Stärke das Neuron “aktiv” wird. Der schichtweise Aufbau und die Aktivierungsfunktionen machen es möglich, dass das neuronale Netz nichtlineare Zusammenhänge darstellen kann [4].

Im Folgenden wird kurz auf die Berechnungsweise eines neuronalen Netzes eingegangen. Sei \mathbf{y}^{l-1} der Ausgabevektor einer bestimmten Schicht $l-1$, so ist dies gleichzeitig der Eingabevektor \mathbf{x}^l für ein Neuron j der nachfolgenden Schicht l , da alle Neuronen einer Schicht mit jedem Neuron der folgenden Schicht verbunden sind. Dementsprechend gilt $\mathbf{y}^{l-1} = \mathbf{x}^l$. Die Ausgabe y_j^l des Neurons berechnet sich nun nach Gleichung (19). Die Ausgabewerte $y_i^{l-1} = x_i^l$ der vorgeschalteten Neuronen i werden mit den jeweiligen Gewichten w_{ji}^{l-1} multipliziert. Die Summe dieser gewichteten Eingaben wird durch die Aktivierungsfunktion α_j^l des betrachteten Neurons ausgewertet. Das Ergebnis ist schließlich die Ausgabe des Neurons j der Schicht l [12].

$$y_j^l = \alpha_j^l \left(\sum_i w_{ji}^{l-1} x_j^{l-1} \right) \quad (19)$$

Das Ziel ist es, die Gewichte w_{ji}^l , die freien Parameter, so zu bestimmen, dass der Zusammenhang zwischen Eingabe und Ausgabe des Netzes möglichst gut wiedergegeben wird. Die Bestimmung dieser Parameter werden im Bereich *Machine Learning* als Training bezeichnet. Neuronale Netze gehören zu den Algorithmen des überwachten Lernens. Dies bedeutet, dass das Netz mit Eingabe- und den dazugehörigen Ausgabe-Werten beaufschlagt wird. Das Training ist ein Optimierungsproblem einer Fehlerfunktion. Das neuronale Netz berechnet mit Hilfe der Eingabewerte einen Ausgabewert und vergleicht diesen mit dem vorgegebenen “wahren” Ausgabewert. Die Fehlerfunktion kann nun z.B. als Summe der Fehlerquadrate zwischen dem berechneten und dem “wahren” Ausgabewert für alle Trainingsdaten definiert werden. Ziel ist es nun diese Fehlerfunktion zu minimieren. Dies geschieht durch Anpassung der Gewichte zwischen den Neuronen. Diese Anpassung kann z.B. mit dem *error backpropagation*-Verfahren durchgeführt werden [4].

3 Datengenerierung

Ziel ist es ein Subgridskalen-Modell zu entwickeln, welches den Stoffübergang an komplexen, sich verändernden Blasenformen in Simulationen mit nicht aufgelöster Konzentrationsgrenzschicht korrigieren kann. Dieses Modell wird mit dem Ansatz neuronaler Netze erstellt. Dafür sind viele Trainingsdaten nötig, die die Bandbreite an auftretenden Feature-Label-Kombinationen abbilden sollen. Zuerst ist der Parameterbereich zu bestimmen, für den das Subgridskalen-Modell einsetzbar sein soll. Anschließend sind Trainingsdaten, die den bestimmten Parameterbereich abdecken, zu generieren. Um den Simulationsaufwand gering zu halten, werden die Trainingsdaten an einem einfachen Ersatzproblem generiert. Außerdem soll das Modell generalisieren können. Es ist nicht möglich jede erdenkliche Blasenform als Trainingsdaten zu hinterlegen, sondern diese durch das Ersatzproblem abzubilden. Es ist nützlich eine Datenreduktion durchzuführen, um die Datenmenge zu verkleinern und redundante Daten herauszufiltern.

3.1 Ersatzproblem

Für die Berechnung des Subgridskalen-Modells werden Trainingsdaten benötigt. Um diese zu berechnen, wird das Problem einer aufsteigenden Gasblase in ein einfacheres Ersatzproblem überführt. Das spätere Subgridskalen-Modell soll verallgemeinern können. Dies bedeutet, dass das Modell, obwohl die Trainingsdaten an einem einfachen Problem generiert werden, auch für andere Fälle (z.B. komplexere Geometrien) anwendbar sein soll. Außerdem verringt das Ersatzproblem den Simulationsaufwand. Dies ist notwendig, um den Parameterbereich mit Simulationen abdecken zu können.

Das Ersatzmodell ist einphasig. Es wird nur die Flüssigphase Ω des Problems betrachtet. Eine weitere Vereinfachung ist, dass nicht eine sich bewegende Blase betrachtet wird, sondern die inverse Sichtweise. Es wird von einer ruhenden Blase ausgegangen, welche angeströmt wird. Dabei wird die Anströmgeschwindigkeit konstant gehalten. In der Natur verändern Blasen beim Aufstieg ihre Form. Dies wird in dem Ersatzproblem nicht berücksichtigt. Es wird eine feste Blasengeometrie simuliert, diese ist sphärisch. Außerdem wird die Vereinfachung getroffen, dass sich das Volumen der Blase nicht verändert, trotz Diffusion des gasförmigen Stoffes in das Fluid. Es wird eine inkompressible Strömung angenommen, d.h. die Dichte ρ der Flüssigkeit ist konstant. Weiterhin wird die Annahme getroffen, dass die drei Spezies A, B und P passiv sind, d.h. sie ändern nicht die Stoffwerte der Flüssigkeit. Für Spezies A wird weiterhin angenommen, dass die Konzentration an der Phasengrenzfläche immer konstant bleibt.

3.2 Numerisches Setup

Im folgenden Abschnitt wird auf die numerische Umsetzung dieses Ersatzproblems eingegangen. Als Simulationssoftware wird OpenFOAM in der Version v1912 benutzt. Als Solver kommt eine angepasste Variante des *pimpleFoam*-Lösers zum Einsatz. Dieser liegt bereits am Institut vor. Um die für die Auswertung benötigten Daten zu extrahieren, wird ein am Institut entwickeltes Utility benutzt. Die Auswertung der Daten und die Erstellung von Graphen erfolgt mit *Jupyter Notebooks*.

Gitter

Die Gitter für die Simulationen werden mit Hilfe eines am Institut geschriebenen Templates erstellt. Da die Blasenform sphärisch ist, kann die Rotationssymmetrie in Umfangsrichtung (senkrecht zur Anströmung) ausgenutzt werden, um die Anzahl an Gitterzellen zu verringern. Das Berechnungsgebiet ist keilförmig und in Umfangsrichtung nur eine Zellschicht dick. Die in OpenFOAM dazugehörige Randbedingung für die Seitenfläche des Keils nennt sich "wedge". Das Gitter ist demnach dreidimensional, wodurch dreidimensionale Geometrieeffekte berücksichtigt werden, wie eine größere Oberfläche im Bereich $\Theta \approx 90^\circ$ im Vergleich zu $\Theta \approx 0^\circ$. Strömungsdynamische Effekte sind hingegen nur zweidimensional. In Abbildung 10 ist eine zweidimensionale Skizze einer Seitenfläche des Berechnungsgebietes zu sehen. Die Zellreihen sind tangential zur Grenzfläche Σ ausgerichtet. Die Flächennormalen, der Zellflächen parallel zu Σ , sind dementsprechend parallel zum Normalenvektor der Grenzfläche. In Abbildung 8 sind die ersten Zellreihen skizziert. Das Template bietet die Möglichkeit die Anzahl der Zellspalten anzugeben, was die Anzahl an Zellen tangential zur Grenzfläche festlegt. Die Angabe bezieht sich auf den Bereich $0^\circ \leq \Theta \leq 90^\circ$. Im unteren Teil der Blase ($90^\circ \leq \Theta \leq 180^\circ$) liegen dreimal so viele Zellen vor, um dieses Gebiet feiner aufzulösen. In diesem Gebiet liegen höhere Konzentrationsgradienten in tangentialer Richtung vor. Außerdem sind die Zellen in den jeweiligen Bereichen nicht gleichmäßig verteilt, sondern werden im Bereich $\Theta = 90^\circ$ dichter. In dem Template kann ebenfalls die Anzahl an Zellreihen, also die Anzahl an Zellen in Normalenrichtung zur Grenzfläche, festgelegt werden. Der äußere Rand des Gitters ($\partial\Omega$) ist ein Kugelausschnitt mit Mittelpunkt identisch zum Mittelpunkt der sphärischen Blase. Dies vermindert Verzerrungen der Zellen, welche prismenförmig sind.

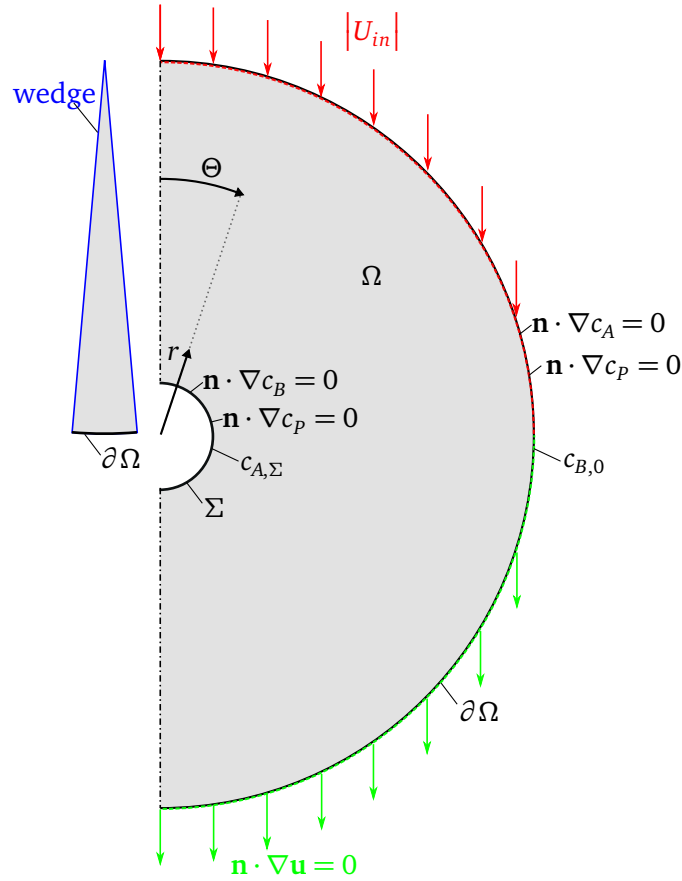


Abbildung 10: Skizze des Berechnungsgebietes mit Randbedingungen.

Randbedingungen

Für die Simulation werden Randbedingungen für die Geschwindigkeit, den Druck und die Stoffverteilungen der drei Spezies benötigt. Tabelle 4 enthält eine Übersicht über die Randbedingungen.

Grenzfläche Σ

Die Grenzfläche Σ stellt die Phasengrenze zwischen der Gasblase und der Flüssigkeit dar. Die Randbedingung für die Geschwindigkeit muss dies physikalisch abbilden. In OpenFOAM wird dafür die Randbedingung “slip” gewählt. Es werden keine Scherkräfte übertragen. Für den Druck, Spezies B und P werden an der Grenzfläche homogene Neumann-Randbedingungen gesetzt. In OpenFOAM wird dies durch die “zeroGradient” Option getan. Der Spezies A wird auf der Grenzfläche der feste Wert $c_{A,\Sigma}$ zugewiesen.

Äußerer Rand $\partial\Omega$

Am äußeren Rand $\partial\Omega$ werden für den Druck, die Spezies A und die Spezies P homogene Neumann-Randbedingungen vorgegeben. Dies wird wieder per “zeroGradient” Option umgesetzt. Wie in Abbildung 10 ersichtlich, wird im oberen Teil ($\Theta \leq 90^\circ$) der äußeren Randfläche $\partial\Omega$ eine Einlassgeschwindigkeit U_{in} vorgegeben. Deren Richtung ist festgelegt, ihr Betrag wird durch die Vorgabe der Reynolds-Zahl bedingt, vergleiche Tabelle 2 und Tabelle 6. Im unteren Teil ($\Theta > 90^\circ$) entspricht die Randbedingung einer Ausflussbedingung, was bedeutet, dass der Gradient der Geschwindigkeit in Normalenrichtung verschwindet. In OpenFOAM ist solch eine Randbedingung als “inletOutlet” verfügbar. Für die Spezies B wird ebenfalls eine “inletOutlet” Randbedingung verwendet. In Einflussgebieten wird die Konzentration über den vorgegebenen Wert $c_{B,0}$ festgelegt.

Seitenflächen

Die Seitenflächen in Umfangsrichtung benötigen aufgrund der Rotationssymmetrie eine Symmetrie-Randbedingung. Da das Berechnungsgebiet einen dreidimensionalen Keil darstellt, wird dies in OpenFOAM durch die Randbedingung “wedge” erfüllt. Diese Symmetrie betrifft alle Größen (Geschwindigkeit, Druck und alle drei Spezies).

Ort	Geschwindigkeit	Druck	Spezies A	Spezies B	Spezies P
Σ	slip	zeroGradient	fixedValue	zeroGradient	zeroGradient
$\partial\Omega$	inletOutlet	zeroGradient	zeroGradient	inletOutlet	zeroGradient
Seitenflächen	wedge	wedge	wedge	wedge	wedge

Tabelle 4: OpenFOAM Randbedingungen für Geschwindigkeit, Druck, Spezies A, B und P.

Anfangsbedingungen

Da es sich um transiente Simulationen handelt, werden neben den Randbedingungen auch Anfangsbedingungen zum Zeitpunkt $t = 0$ benötigt. Diese werden analog zu den Randbedingungen für die Geschwindigkeit, den Druck und die drei Spezies A, B und P gebraucht und sind in Tabelle 5 zusammengefasst.

Bei $t = 0$ liegt eine homogene Geschwindigkeitsverteilung vor mit Betrag und Richtung der vorgegebenen Randbedingung \mathbf{U}_{in} . Für den Druck sind nicht die Absolutwerte entscheidend, da der Referenz- bzw. der Umgebungsdruck frei gewählt werden kann. Von Bedeutung sind die sich einstellenden Druckdifferenzen im Berechnungsgebiet. Dementsprechend sind die berechneten Druckwerte immer Angaben relativ zum Referenzdruck. Die Angabe $p = 0 \text{ kg/ms}^2$ für eine beliebige Stelle innerhalb des Rechenggebietes bedeutet, dass an dieser Stelle der Referenzdruck vorliegt. Als Startbedingung liegt im kompletten Berechnungsgebiet der Referenzdruck vor, also gilt $p(\Omega, t = 0) = 0 \text{ kg/ms}^2$. Im Berechnungsgebiet liegt zu Beginn der Simulation noch keine Spezies A vor. Diese muss zuerst von der Grenzfläche Σ aus in das Berechnungsgebiet diffundieren. Die Anfangsbedingung ergibt sich also zu $c_A(\Omega, t = 0) = 0 \text{ mol/m}^3$. Bei $t = 0$ ist Spezies B homogen in der Flüssigkeit verteilt und hat überall den vorgegebenen Randwert, dementsprechend gilt $c_B(\Omega, t = 0) = c_{B,0}$. Zu Beginn der Simulation liegt noch keine Spezies P vor, dementsprechend gilt $c_P(\Omega, t = 0) = 0 \text{ mol/m}^3$ im gesamten Berechnungsgebiet.

Geschwindigkeit	Druck	Spezies A	Spezies B	Spezies P
$\mathbf{U}(\Omega, t = 0) = \mathbf{U}_{in}$	$p(\Omega, t = 0) = 0 \frac{\text{kg}}{\text{m s}^2}$	$c_A(\Omega, t = 0) = 0 \frac{\text{mol}}{\text{m}^3}$	$c_B(\Omega, t = 0) = c_{B,0}$	$c_P(\Omega, t = 0) = 0 \frac{\text{mol}}{\text{m}^3}$

Tabelle 5: Anfangsbedingungen für Geschwindigkeit, Druck, Spezies A, B und P.

Charakteristische Größen

In Kapitel 2.1 und 2.2 wurden charakteristische Größen verwendet um die strömungsdynamischen und Stofftransport-Gleichungen zu entdimensionalisieren. Als charakteristische Geschwindigkeit \hat{U} wird der Betrag der Einlassgeschwindigkeit \mathbf{U}_{in} gewählt. Für die charakteristische Länge \hat{L} bietet sich der Durchmesser der Blase d_{eq} an. Die Konzentration der Spezies A an der Grenzfläche $c_{A,\Sigma}$ wird als charakteristische Konzentration \hat{c} genommen. Tabelle 6 fasst dies zusammen.

Geschwindigkeit	Länge	Konzentration
$\hat{U} = \mathbf{U}_{in} $	$\hat{L} = d_{eq}$	$\hat{c} = c_{A,\Sigma}$

Tabelle 6: Charakteristische Größen.

3.3 Parameterbereich

Es gilt einer Reihe von Größen Werte zu zuweisen. Dies betrifft sowohl Konstanten, die in die Grundgleichungen der Strömungsdynamik bzw. des Stofftransports eingehen, als auch Randwertbedingungen. Diese sind in Tabelle 7 zusammengefasst. Diese Größen stehen in direkter Beziehung zu den dimensionslosen Kennzahlen.

In [11] werden experimentelle und numerische Daten verglichen. Für die Experimente wird die Kombination aus Wasser als Flüssigkeit und eine Gasblase aus Sauerstoff verwendet. Unter Standardbedingungen (20°) hat Wasser eine kinematische Viskosität von $\nu \approx 10^{-6} \text{ m}^2/\text{s}$ und eine Dichte von $\rho \approx 1000 \text{ kg/m}^3$ [14]. Diese Werte werden für die Simulationen in dieser Arbeit übernommen. Der Durchmesser der Blase wird zu $d_{eq} = 0.002 \text{ m}$ festgelegt. Die Einlassgeschwindigkeit kann nun über die Festlegung der Re -Zahl bestimmt werden. Die Re soll im Bereich $[10, 250]$ liegen, damit lässt sich der Bereich aus [11] und [3] abdecken. Der Diffusionskoeffizient D wird durch die Wahl der Schmidt-Zahl Sc festgelegt. Wenn man den Diffusionskoeffizient von Sauerstoff in Wasser anlegt ($D_{O_2} = 2 \cdot 10^{-9} \text{ m}^2/\text{s}$) ergibt sich $Sc = 500$. Der Pe -Bereich ergibt sich damit zu $[5000, 125000]$. Die Konzentration von Spezies A an der Grenzfläche wird zu $c_{A,\Sigma} = 1 \text{ mol/m}^3$ festgesetzt. In [11] werden Konzentrationen von Spezies B $\tilde{c}_{B,0}$ in der Flüssigkeit im Bereich von $[0.9, 3.46]$ betrachtet, diese Werte sind relativ zur Konzentration von Spezies A an der Grenzfläche. Dieser Bereich wird für diese Arbeit etwas vergrößert auf $[0.2, 5]$. Die Reaktionsgeschwindigkeitskonstante k wird durch die Damköhler-Zahl festgelegt. Diese soll

in einem Bereich von $[0, 0.3]$ variiert werden. Als Endzeit für die Simulationen wird die dimensionslose Zeit $\tilde{t} = 5$ gewählt, sodass sich ein stationärer Zustand einstellt. Für die jeweilige Re -Zahl muss diese in eine Zeit $t = \tilde{t} d_{eq}^2 / (Re \nu)$ umgerechnet werden.

kinematische Viskosität	ν
Durchmesser	d_{eq}
Einflussgeschwindigkeit	$ \mathbf{U}_{in} $
Reaktionsgeschwindigkeitskonstante	k
Diffusionskoeffizient	D
Konzentration von Spezies A an der Grenzfläche	$c_{A,\Sigma}$
Konzentration von Spezies B am äußeren Rand	$c_{B,0}$

Tabelle 7: Zu definierende Größen.

3.4 Größe des Berechnungsgebiets

Wie in Abbildung 10 zu erkennen ist, ist der äußere Rand des Berechnungsgebietes ein Halbkreis mit dem Durchmesser $d_{\partial\Omega}$. Randeffekte durch ein zu kleines Berechnungsgebiet, die auf den grenzflächennahen Bereich einwirken, sollen ausgeschlossen werden. Aber gleichzeitig soll das Berechnungsgebiet nicht unnötig groß gewählt werden. Um dies zu untersuchen, wird im Folgenden das Verhältnis von $d_{\partial\Omega}$ zu d_{eq} variiert. Da bei großen Schmidt-Zahlen Sc der Stofftransport konvektionsdominiert ist, genügt es zu zeigen, dass der Wandeinfluss auf das Geschwindigkeitsfeld vernachlässigbar ist, um daraus schließen zu können, dass auch für den Stofftransport der Wandeinfluss vernachlässigbar ist. Daraus folgt, dass der Einfluss auf die tangentielle Geschwindigkeit U_τ an der Grenzfläche Σ betrachtet wird. Es werden nur die strömungsdynamischen Gleichungen gelöst. Aufgrund der größeren Grenzschichtdicke kann eine gröbere räumliche Auflösung verwendet werden, als wenn der Stofftransport mit berücksichtigt wird. Der Simulationsaufwand wird geringer. Bei kleineren Reynolds-Zahlen Re ist die dynamische Grenzschichtdicke größer als bei größeren Re . Ein zu kleines Berechnungsgebiet führt zuerst bei kleinen Re zu den unerwünschten Randeffekten, weswegen diese Untersuchung bei kleinen Re durchgeführt wird. Zu beachten ist bei den Simulationsergebnissen dieses Abschnittes, dass diese an einer ellipsoiden Blasenform (Seitenverhältnis $\chi = 3$) gewonnen wurden. Der Hintergrund ist, dass eine eventuelle Ausweitung der Trainingsdaten auf ellipsoide Blasenformen möglich ist. Die Ergebnisse für die Bestimmung von $d_{\partial\Omega}$ sind jedoch auf sphärische Blasen anwendbar, da bei gleichem $d_{\partial\Omega}$ der Wandeinfluss bei ellipsoiden Blasen eher zum Tragen kommt als bei sphärischen Blasen.

Die Reynolds-Zahl wird im Bereich $Re = 1$ bis $Re = 100$ variiert. Die Variation der Größe des Berechnungsgebietes, repräsentiert durch das Verhältnis $\frac{d_{\partial\Omega}}{d_{eq}}$, geschieht im Bereich $\frac{d_{\partial\Omega}}{d_{eq}} = 3$ bis $\frac{d_{\partial\Omega}}{d_{eq}} = 10$.

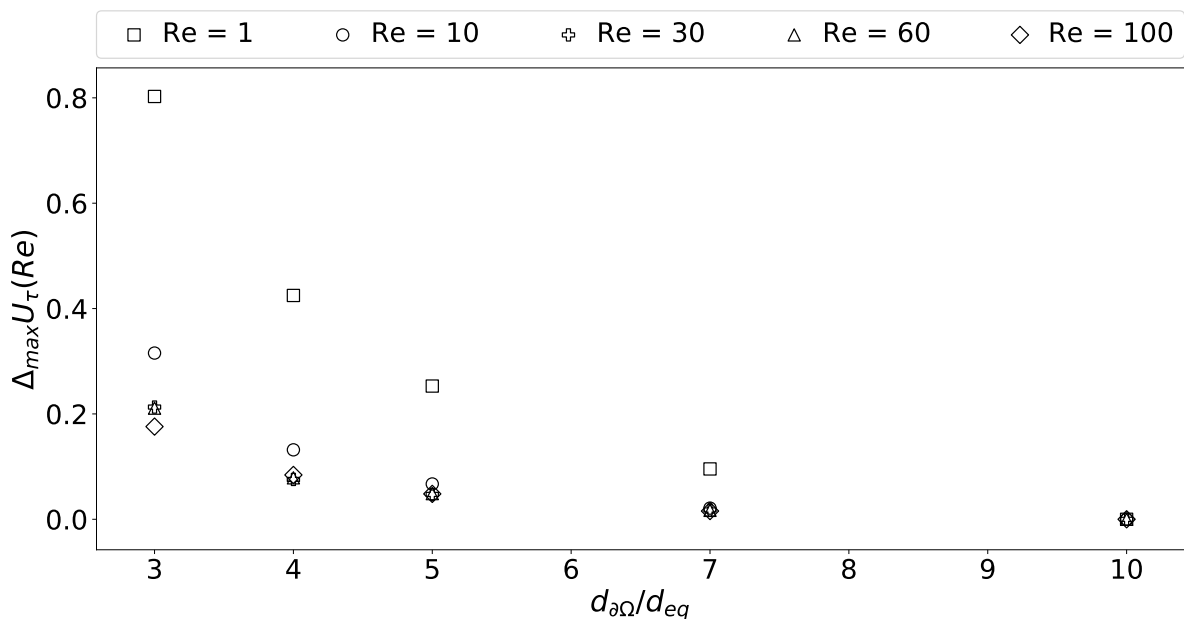


Abbildung 11: Normierte maximale Abweichung $\Delta_{max} U_\tau(Re)$, siehe Gleichung (20).

In Abbildung 11 wird die maximale Abweichung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche für unterschiedlich große Berechnungsgebiete und unterschiedliche Reynolds-Zahlen betrachtet. Die maximale Abweichung $\Delta_{\max} U_\tau(Re)$ berechnet sich nach Gleichung (20). Hierbei fließt eine Normierung durch die maximale Geschwindigkeit auf dem größten Berechnungsgebiet bei der jeweiligen Reynolds-Zahl ein. Dies dient dazu die unterschiedlichen Reynolds-Zahlen vergleichbar zu machen.

$$\Delta_{\max} U_\tau(Re) = \frac{\max \left[\left| U_\tau(d_{\partial\Omega}/d_{eq} = 10, Re) - U_\tau(d_{\partial\Omega}/d_{eq}, Re) \right| \right]}{\max \left[\left| U_\tau(d_{\partial\Omega}/d_{eq} = 10, Re) \right| \right]} \quad (20)$$

Es ist zu erkennen, dass mit größer werdendem Berechnungsgebiet die maximale Abweichung für alle Reynolds-Zahlen abnimmt. Außerdem nimmt die maximale Abweichung für alle $\frac{d_{\partial\Omega}}{d_{eq}}$ mit steigender Reynolds-Zahl ab. Dies bestätigt die Annahme für diese Untersuchung kleine Reynolds-Zahlen zu betrachten. Für $\frac{d_{\partial\Omega}}{d_{eq}} \leq 5$ treten bei allen Re deutliche Abweichungen ($> 4\%$) hinsichtlich $\frac{d_{\partial\Omega}}{d_{eq}} = 10$ auf. Für $Re = 1$ sind diese nochmals deutlich größer als für die anderen Re .

Gewählt wird $\frac{d_{\partial\Omega}}{d_{eq}} = 7$, da bei dieser Größe alle betrachteten Reynolds-Zahlen $Re > 1$ vernachlässigbare Abweichungen ($< 4\%$) zur Referenzgröße $\frac{d_{\partial\Omega}}{d_{eq}} = 10$ aufweisen, siehe Tabelle 8.

Reynolds-Zahl	Abweichung
1	9.6%
10	2.1%
30	1.8%
60	1.7%
100	1.6%

Tabelle 8: Abweichungen für $\frac{d_{\partial\Omega}}{d_{eq}} = 7$.

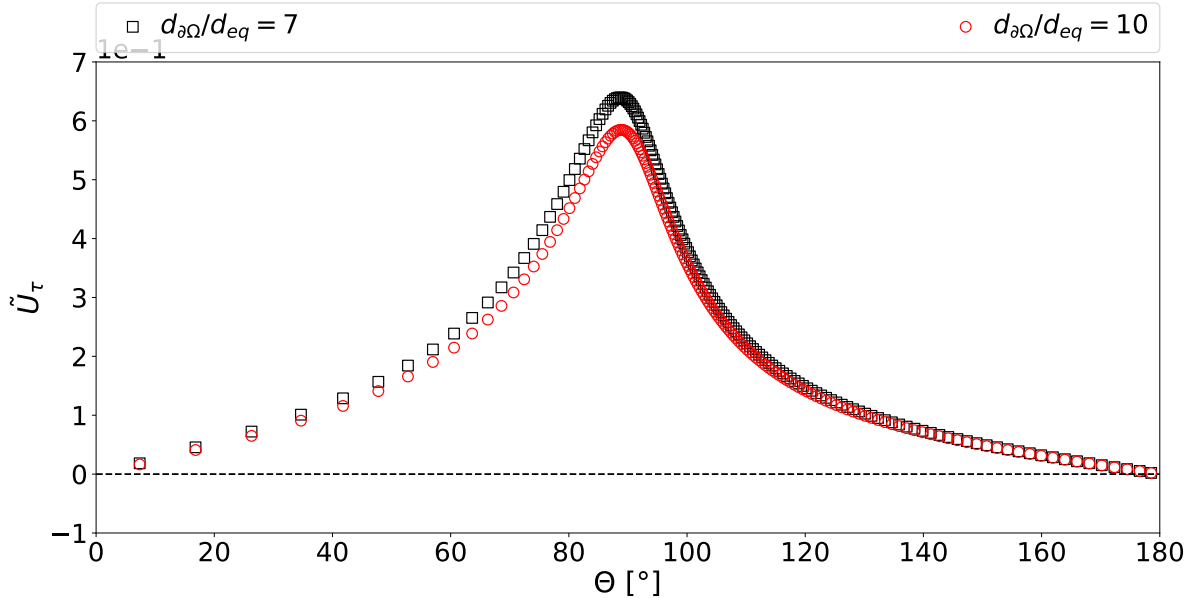


Abbildung 12: Verteilung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche in Abhängigkeit vom Polarwinkel Θ bei $Re = 1$ für unterschiedlich große Berechnungsgebiete. Die gestrichelte Linie ist die Nulllinie.

Nachfolgend ist in den Abbildungen 12 und 13 die lokale Verteilung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche dargestellt. Die Geschwindigkeit ist mit der jeweiligen Einlassgeschwindigkeit entdimensionalisiert. Abbildung 12

gilt für $Re = 1$. Es sind deutliche Unterschiede in der Geschwindigkeitsverteilung im Vergleich zum größten Berechnungsgebiet zu erkennen. Die maximale Differenz tritt bei $\Theta \approx 90^\circ$ auf. Die gestrichelte Linie stellt die Nulllinie dar. Es ist zu erkennen, dass $\tilde{U}_\tau > 0$ gilt. Es liegen keine Rezirkulationsgebiete vor.

Abbildung 13 gilt für den Fall $Re = 100$. Die Geschwindigkeitsverteilungen nähern sich deutlich mehr der Verteilung für die Referenzgröße an als bei dem vorherigen Fall. Die maximale Differenz liegt ebenso bei $\Theta \approx 90^\circ$. In diesem Fall wird im unteren Teil der Blase ($\Theta > 90^\circ$) die Tangentialgeschwindigkeit negativ. Dies deutet auf ein Rezirkulationsgebiet hin.

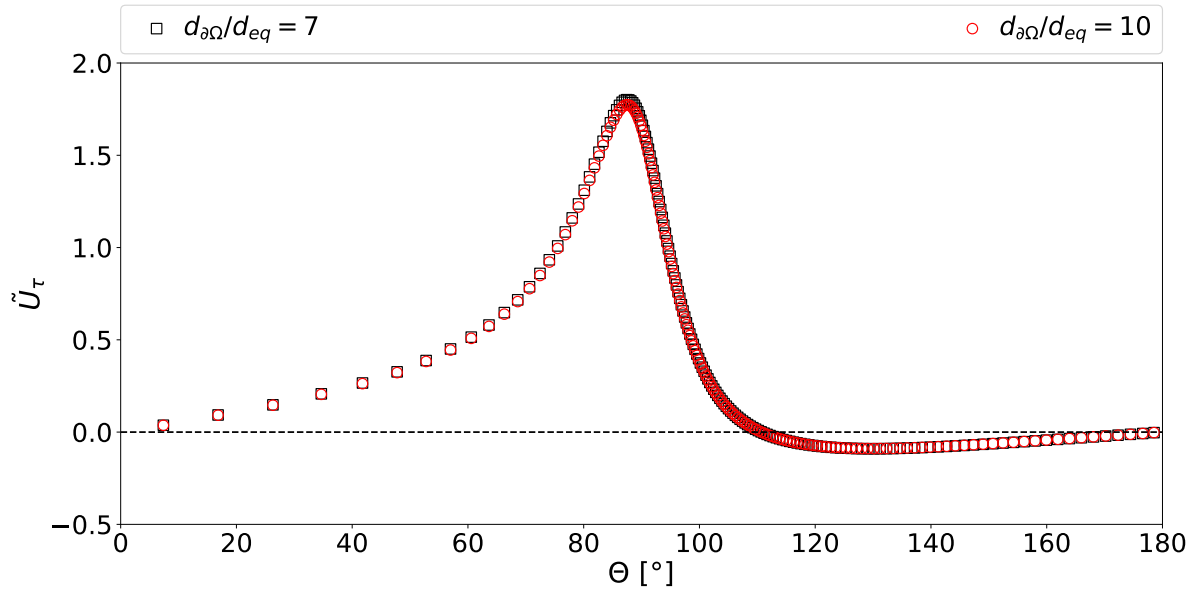


Abbildung 13: Verteilung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche in Abhängigkeit vom Polarwinkel Θ bei $Re = 100$ für unterschiedlich große Berechnungsgebiete. Die gestrichelte Linie ist die Nulllinie.

3.5 Räumliche Auflösung

Für die Durchführung der Simulationen mit dem Ziel der Auswertung des Stofftransports muss ein dafür geeignetes Gitter gewählt werden. Die räumliche Auflösung muss fein genug sein, um die Konzentrationsgrenzschichten aufzulösen. Der Prozess ein geeignetes Gitter zu finden wird Gitterstudie genannt, dabei wird gezeigt, dass ein feineres Gitter keine Auswirkungen auf die betrachteten Größen hat. Der Extremfall, also die kleinste Grenzschicht für Spezies A, tritt bei der größten Péclet-Zahl und der stärksten Reaktion, also der größten Damköhler-Zahl, auf. Dies bedeutet, es wird der Fall betrachtet mit $Re = 250$ und $Sc = 500$, woraus $Pe = 125000$ folgt. Außerdem gilt $Da = 0.3$ und $c_{B,0} = 5$, da die Konzentration von Spezies B ebenfalls in den Reaktionsterm der Stofftransportgleichung eingeht. Betrachtet werden die lokalen Sherwood-Zahlen über der Blasengrenzfläche, die globale Sherwood-Zahl und der Konzentrationsverlauf in radialer Richtung für einen bestimmten, frei wählbaren Winkel. Die Sherwood-Zahl ist eine dimensionslose Kennzahl, die den Stoffübergang beschreibt. Die lokale Sherwood-Zahl ist definiert als die entdimensionalisierte Normalenableitung der Konzentration von Spezies A an der Grenzfläche (siehe Gleichung (21)), aus [10].

$$Sh_{loc} = \frac{\left| \partial_n c_A \right|_{\Sigma} \cdot d_{eq}}{c_{A,\Sigma}} \quad (21)$$

Die globale Sherwood-Zahl ergibt sich aus der Mittelung der lokalen Sherwood-Zahlen über der Grenzfläche (siehe Gleichung 22).

$$Sh = \frac{1}{A_{\Sigma}} \int_{\Sigma} Sh_{loc} dA \quad (22)$$

Es ist die Besonderheit des Gitter-Templates zu beachten, wonach im Bereich $\Theta \leq 90^\circ$ ein Viertel und im Bereich $\Theta > 90^\circ$ drei Viertel der tangentialen Zellen sind (siehe Kapitel 3.2). Betrachtet werden im Folgenden drei verschiedene Gitter. Das Größte hat 200 Zellen in tangentialer und 70 Zellen in radialer Richtung. Das zweite Gitter besteht aus 280 tangentialen und 100 radialen Zellen und das feinste Gitter hat 400 Zellen in tangentialer und 130 Zellen in radialer Richtung. Außerdem wird ein radiales Grading von 900 gewählt.

In Abbildung 14 sind die Verläufe der lokalen Sherwood-Zahl über dem Polarwinkel Θ (oben) und der Konzentrationsverlauf der Spezies A in radialer Richtung (unten) dargestellt. Es ist zu beachten, dass der Konzentrationsverlauf bei ungefähr 92° abgebildet ist, da alle drei Gitter eine unterschiedliche tangentiale Auflösung haben, aber im Bereich $\Theta \approx 90^\circ$ die Verteilung der tangentialen Zellen am dichtesten ist. Die Werte sind 92.10° (grobes Gitter), 91.93° (mittleres Gitter) und 91.95° (feines Gitter). Wie zu erkennen ist, gibt es kleinere Abweichungen der lokalen Sherwood-Zahl des größten Gitters von den beiden Anderen bei kleinen Winkeln Θ . Für $\Theta > 80^\circ$ sind keine Abweichungen der drei Gitter erkennbar. Für den Konzentrationsverlauf in Normalenrichtung ergeben sich keine erkennbaren Unterschiede zwischen den drei Gittervarianten. In Tabelle 9 sind die globalen Sherwood-Zahlen für die drei Varianten aufgelistet. Die Abweichungen sind in Bezug auf die betragsmäßige Größe der Werte gering.

Es wird das mittlere Gitter verwendet, um eine bessere Auflösung bei kleinen Θ zu haben. Das feine Gitter ist ebenso verwendbar, jedoch führt dies zu einer längeren Simulationsdauer.

Gitter	Sherwood-Zahl
200x70	530.17
280x100	531.67
400x130	532.28

Tabelle 9: Globale Sherwood-Zahl.

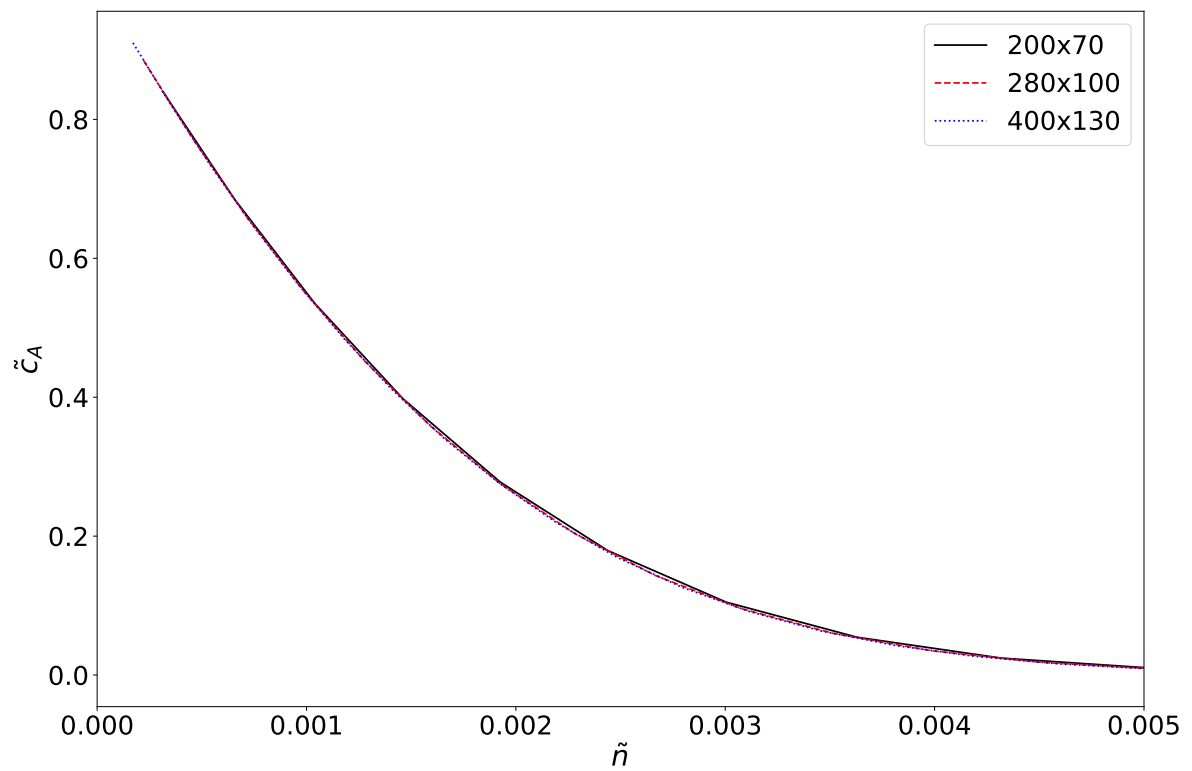
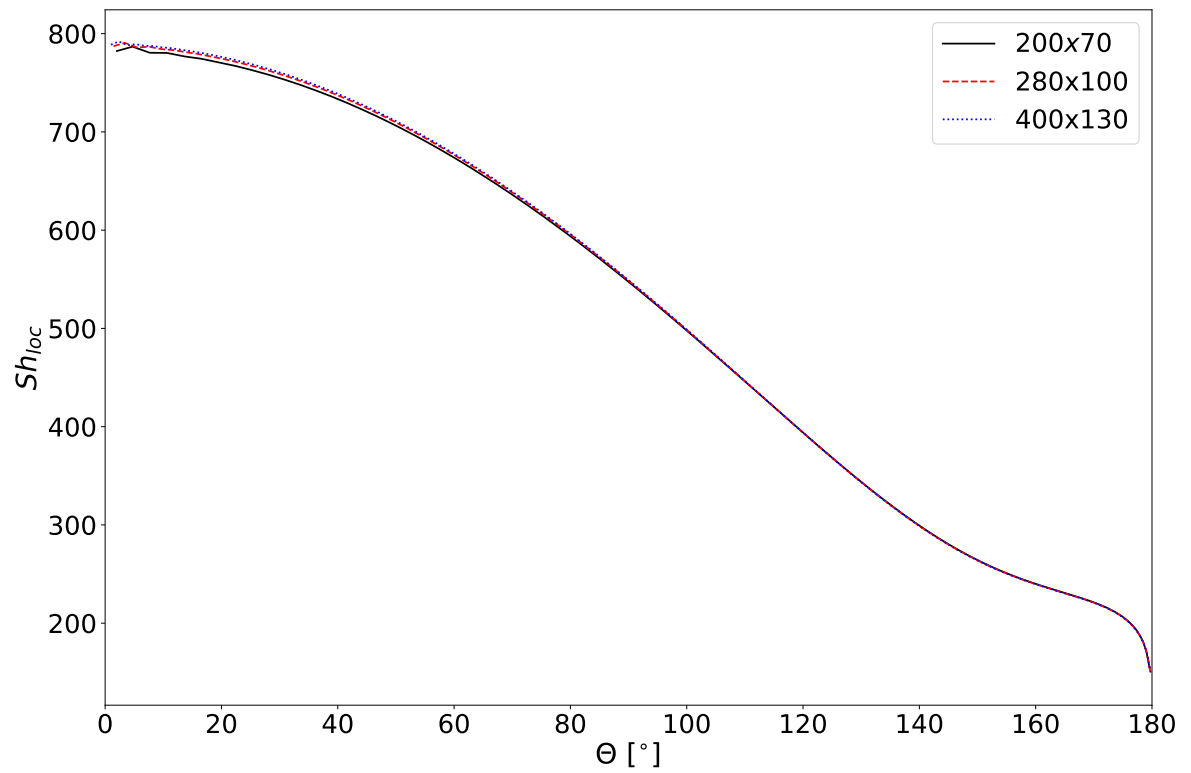


Abbildung 14: Gitterstudie mit unterschiedlich feinen Gittern. Verläufe der lokalen Sherwood-Zahl über Polarwinkel Θ (oben), Verläufe der Konzentration der Spezies A in radialer Richtung für $\Theta \approx 92^\circ$ (unten). ($Re = 250$, $Sc = 500$, $Da = 0.3$, $c_{B,0} = 5$)

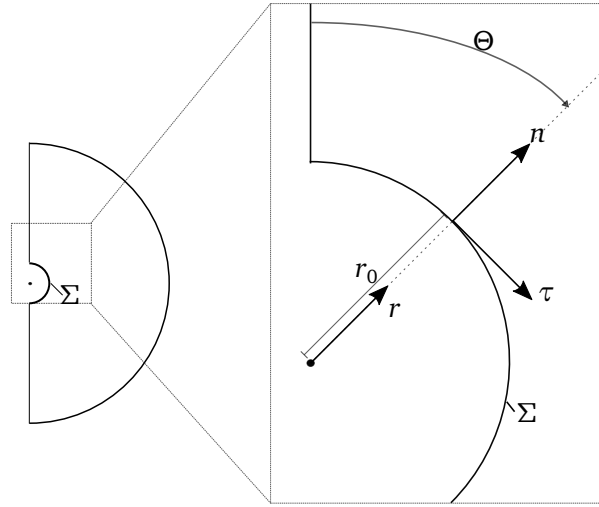


Abbildung 15: Skizze der Koordinatensysteme.

Um den Parameterbereich (siehe 3.3) abzudecken, sind immer noch viele Simulationen nötig, wobei bei den Simulationen auch viele nicht benötigte Daten entstehen. Für die Generierung von Trainingsdaten nicht genutzte Bereiche sind z.B. Gebiete in der Nähe des äußeren Randes $\partial\Omega$, in denen keine Konzentrationsgradienten vorliegen.

Um die Datengenerierung zu beschleunigen und die Menge an berechneten Daten vorab möglichst gering zu halten, wird das Ersatzmodell weiter vereinfacht. Von Interesse für die Trainingsdaten ist der Bereich um die Grenzfläche, in dem das Subgridskalen-Modell später Korrekturen ausführen soll und auch dort sind hauptsächlich die Konzentrationsgrenzschichten der unterschiedlichen Spezies von Interesse. In Anlehnung an das Vorgehen in [1] werden die Konzentrationsgrenzschichten durch ein eindimensionales Modell (in Normalenrichtung zur Grenzfläche Σ) berechnet. Dies geschieht durch Vereinfachung der zweidimensionalen Transportgleichungen für die Spezies A, B und P. In Gleichung (23) ist die entdimensionalisierte Transportgleichung in Kugelkoordinaten gegeben, wobei die dritte Dimension durch die Annahme von Symmetrie in Umfangsrichtung schon vernachlässigt wurde.

$$\frac{\partial \tilde{c}_i}{\partial \tilde{t}} = \frac{1}{Pe} \left[\frac{\partial^2 \tilde{c}_i}{\partial \tilde{r}^2} + \frac{2}{\tilde{r}} \frac{\partial \tilde{c}_i}{\partial \tilde{r}} + \frac{1}{\tilde{r}^2} \frac{\partial^2 \tilde{c}_i}{\partial \Theta^2} + \frac{1}{\tilde{r}^2 \tan(\Theta)} \frac{\partial \tilde{c}_i}{\partial \Theta} \right] - \tilde{U}_{\tilde{r}} \frac{\partial \tilde{c}_i}{\partial \tilde{r}} - \tilde{U}_{\Theta} \frac{1}{\tilde{r}} \frac{\partial \tilde{c}_i}{\partial \Theta} \pm Da \tilde{c}_A \tilde{c}_B \quad (23)$$

Diese muss zuerst in das gewünschte, lokale Koordinatensystem (n, τ) überführt werden. Im Gegensatz zu den Kugelkoordinaten, welche ihren Ursprung im Mittelpunkt der Blase haben, hat dieses Koordinatensystem seinen Ursprung auf der Grenzfläche Σ . Die beiden Koordinaten sind die Normale (n) und Tangentiale (τ) zur Grenzfläche. Je nach betrachteter Stelle (bezüglich des Polarwinkels Θ) auf der Grenzfläche ändern sich die Richtungen von n und τ bezogen auf ein inertiales Koordinatensystem (x, y) . Durch einen Vergleich von Kugelkoordinaten (r, Θ) und (n, τ) -Koordinaten lässt sich Gleichung (23) überführen in Gleichung (24).

$$\frac{\partial \tilde{c}_i}{\partial \tilde{t}} = \frac{1}{Pe} \left[\frac{\partial^2 \tilde{c}_i}{\partial \tilde{n}^2} + \frac{2}{\tilde{r}_0 + \tilde{n}} \frac{\partial \tilde{c}_i}{\partial \tilde{n}} + \frac{\partial^2 \tilde{c}_i}{\partial \tilde{\tau}^2} + \frac{1}{(\tilde{r}_0 + \tilde{n}) \tan(\Theta)} \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}} \right] - \tilde{U}_n \frac{\partial \tilde{c}_i}{\partial \tilde{n}} - \tilde{U}_{\tau} \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}} \pm Da \tilde{c}_A \tilde{c}_B \quad (24)$$

Die mathematische Herleitung ist im Anhang zu finden (siehe Anhang A).

Term-Analyse

Das Modell soll so einfach wie möglich gehalten werden. Deswegen wird zunächst die Bedeutung der Terme der rechten Seite untersucht, um vernachlässigbare Terme zu identifizieren. In den Abbildungen 16, 17, 18 und 19 sind die Verläufe der einzelnen Terme in Normalenrichtung dargestellt.

T1	T2	T3	T4	T5	T6	T7	Summe
$\frac{1}{Pe} \frac{\partial^2 \tilde{c}_i}{\partial \tilde{n}^2}$	$\frac{1}{Pe} \frac{2}{\tilde{r}_0 + \tilde{n}} \frac{\partial \tilde{c}_i}{\partial \tilde{n}}$	$\frac{1}{Pe} \frac{\partial^2 \tilde{c}_i}{\partial \tilde{\tau}^2}$	$\frac{1}{Pe} \frac{1}{(\tilde{r}_0 + \tilde{n}) \tan(\Theta)} \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}}$	$-\tilde{U}_n \cdot \frac{\partial \tilde{c}_i}{\partial \tilde{n}}$	$-\tilde{U}_\tau \cdot \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}}$	$\pm Da \cdot \tilde{c}_A \cdot \tilde{c}_B$	$\frac{\partial \tilde{c}_i}{\partial \tilde{t}}$

Tabelle 10: Erläuterung der Legendenbezeichnungen in den Abbildungen 16, 17, 18 und 19.

In Abbildung 16 werden vergleichend die Verläufe für die unterschiedlichen Spezies gegenübergestellt. Wie zu erkennen ist, sind nur die Terme $T1, T5, T6$ und $T7$ signifikant von Null verschieden. Unter Summe versteht sich die Summe aller Terme der rechten Seite von Gleichung (24), also die zeitliche Änderung der Konzentration der jeweiligen Spezies. Für eine genaue Bezeichnung der Terme sei auf Tabelle 10 verwiesen. Abbildung 16 gilt für den Polarwinkel $\Theta = 90^\circ$. Unter anderen Winkeln kann sich die Größe der Terme ändern und andere Terme können an Bedeutung gewinnen. Dies ist in Abbildung 17 für Spezies A dargestellt. Auch hier ist zu erkennen, dass die Terme $T2, T3$ und $T4$ unter keinem Polarwinkel an Bedeutung gewinnen. Wie jedoch ein Vergleich der drei Grafiken in Abbildung 17 zeigt, können sich Größe und Vorzeichen der Terme ändern, so z.B. an Term $T5$ ersichtlich. Dies ist wenig verwunderlich, da in diesen Term die Geschwindigkeit normal zur Grenzfläche eingeht und diese sich mit dem Polarwinkel ändert. Abbildung 18 zeigt die Auswertung der Terme zu unterschiedlichen Zeiten für Spezies A. Wie zu erkennen ist, ist in der unteren Grafik (bei $\tilde{t} = 0.125$) noch kein stationärer Zustand erreicht. Dies ist daran zu erkennen, dass der Summenterm nicht an jeder Stelle verschwindet. Bei $\tilde{t} = 5$, Grafik oben, ist der stationäre Zustand erreicht. In beiden Fällen sind die Terme $T2, T3$ und $T4$ vernachlässigbar gering. In Abbildung 19 wird die Auswirkung unterschiedlicher Péclet-Zahlen miteinander verglichen. An der Größenordnung der Terme ändert sich wenig. Es ist jedoch der Einfluss der Péclet-Zahl auf die Größe des Bereichs (in \tilde{n} -Richtung), indem die Terme sich deutlich von Null unterscheiden, zu erkennen. Je kleiner die Péclet-Zahl desto größer ist dieser Bereich. Zusammenfassend ist festzuhalten, dass die Beiträge der Terme $T2 = \frac{1}{Pe} \frac{2}{\tilde{r}_0 + \tilde{n}} \frac{\partial \tilde{c}_i}{\partial \tilde{n}}$, $T3 = \frac{1}{Pe} \frac{\partial^2 \tilde{c}_i}{\partial \tilde{\tau}^2}$ und $T4 = \frac{1}{Pe} \frac{1}{(\tilde{r}_0 + \tilde{n}) \tan(\Theta)} \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}}$ gegenüber den Restlichen vernachlässigbar klein sind. Diese werden im Folgenden weggelassen und somit ergibt sich Gleichung (25). In Anhang B sind ähnliche Abbildungen für die Spezies B und P zu finden.

$$\frac{\partial \tilde{c}_i}{\partial \tilde{t}} = \frac{1}{Pe} \cdot \frac{\partial^2 \tilde{c}_i}{\partial \tilde{n}^2} - \tilde{U}_n \cdot \frac{\partial \tilde{c}_i}{\partial \tilde{n}} - \tilde{U}_\tau \cdot \frac{\partial \tilde{c}_i}{\partial \tilde{\tau}} \pm Da \cdot \tilde{c}_A \cdot \tilde{c}_B \quad (25)$$

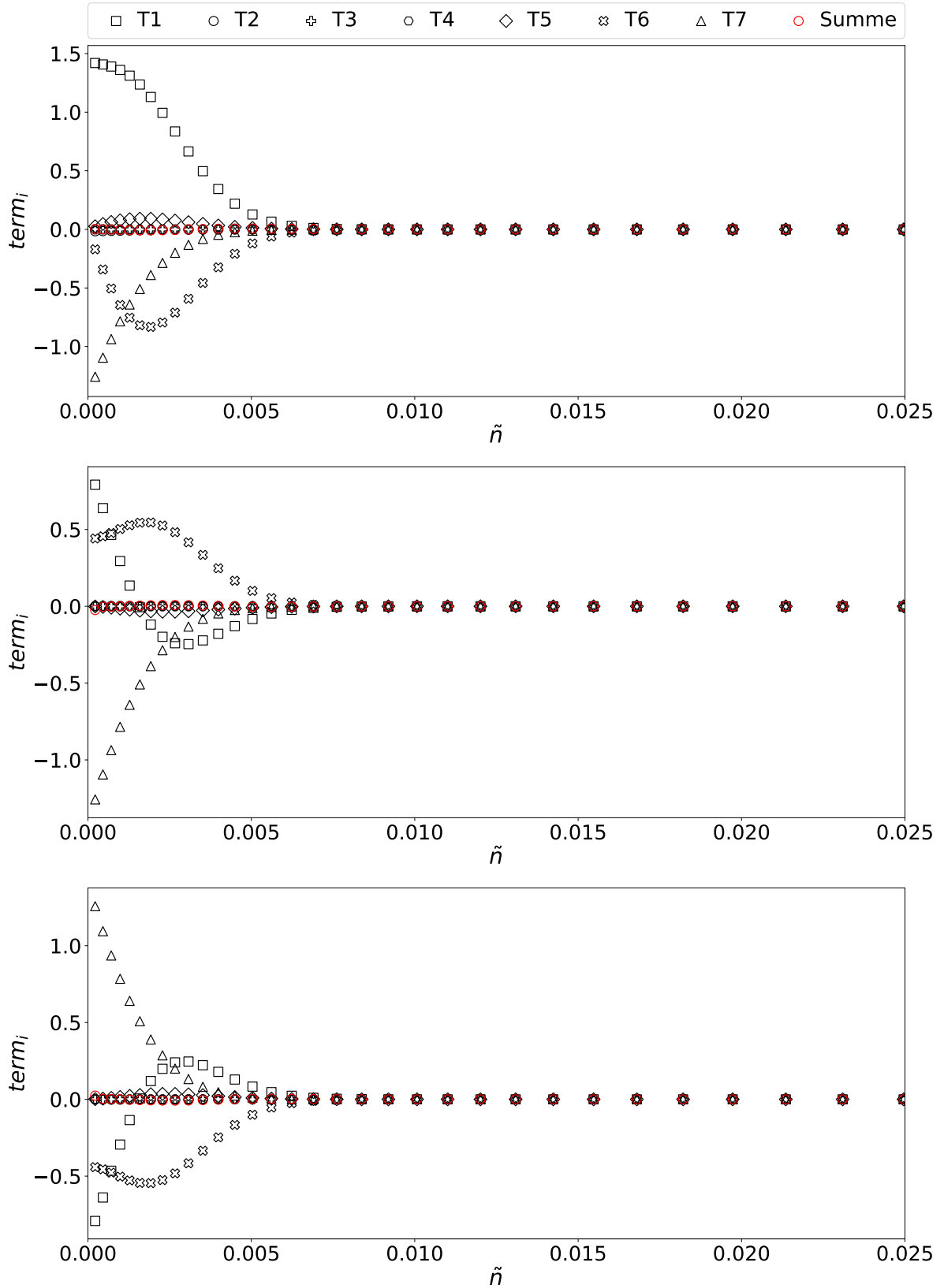


Abbildung 16: Verlauf der Terme der linken Seite aus Gleichung (24) für unterschiedliche Spezies bei $\Theta = 90^\circ$. Oben Spezies A, mitte Spezies B, unten Spezies P ($Pe = 125000$; $Da = 0.3$; $c_{B,0} = 5$).

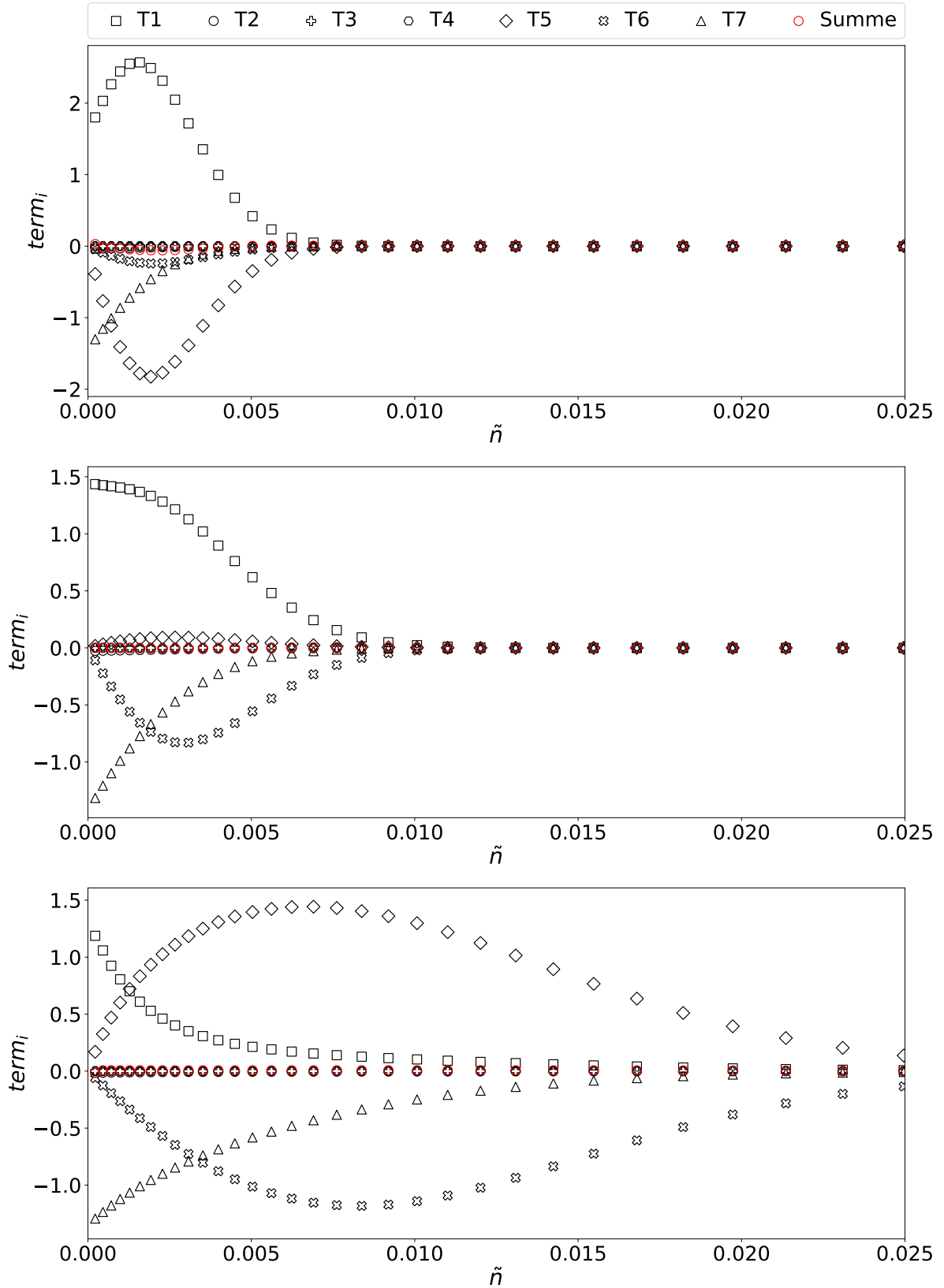


Abbildung 17: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).

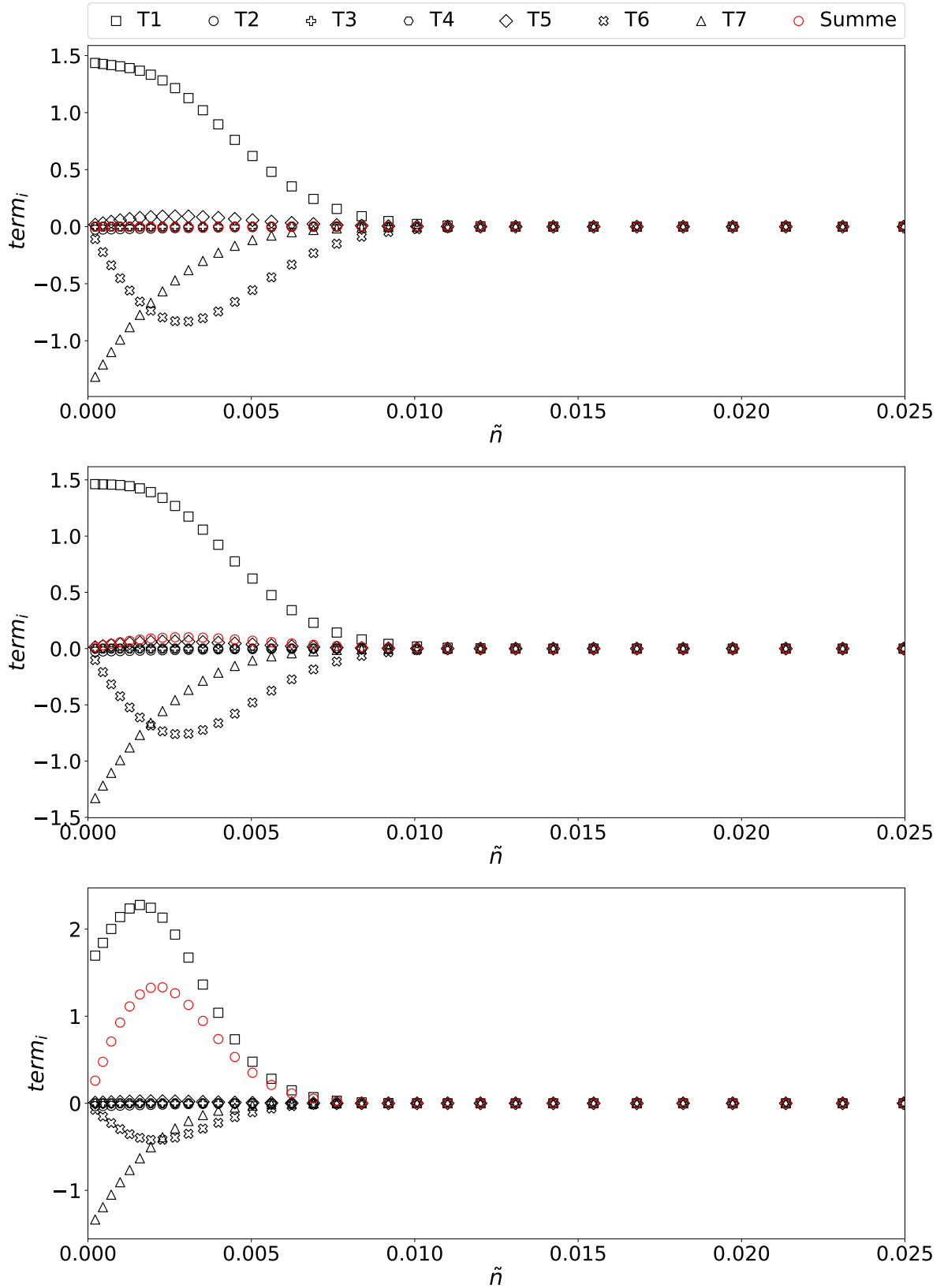


Abbildung 18: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A zu unterschiedlichen Zeiten bei Polarwinkel $\Theta = 90^\circ$. Oben $\tilde{t} = 5$, Mitte $\tilde{t} = 0.375$, unten $\tilde{t} = 0.125$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).

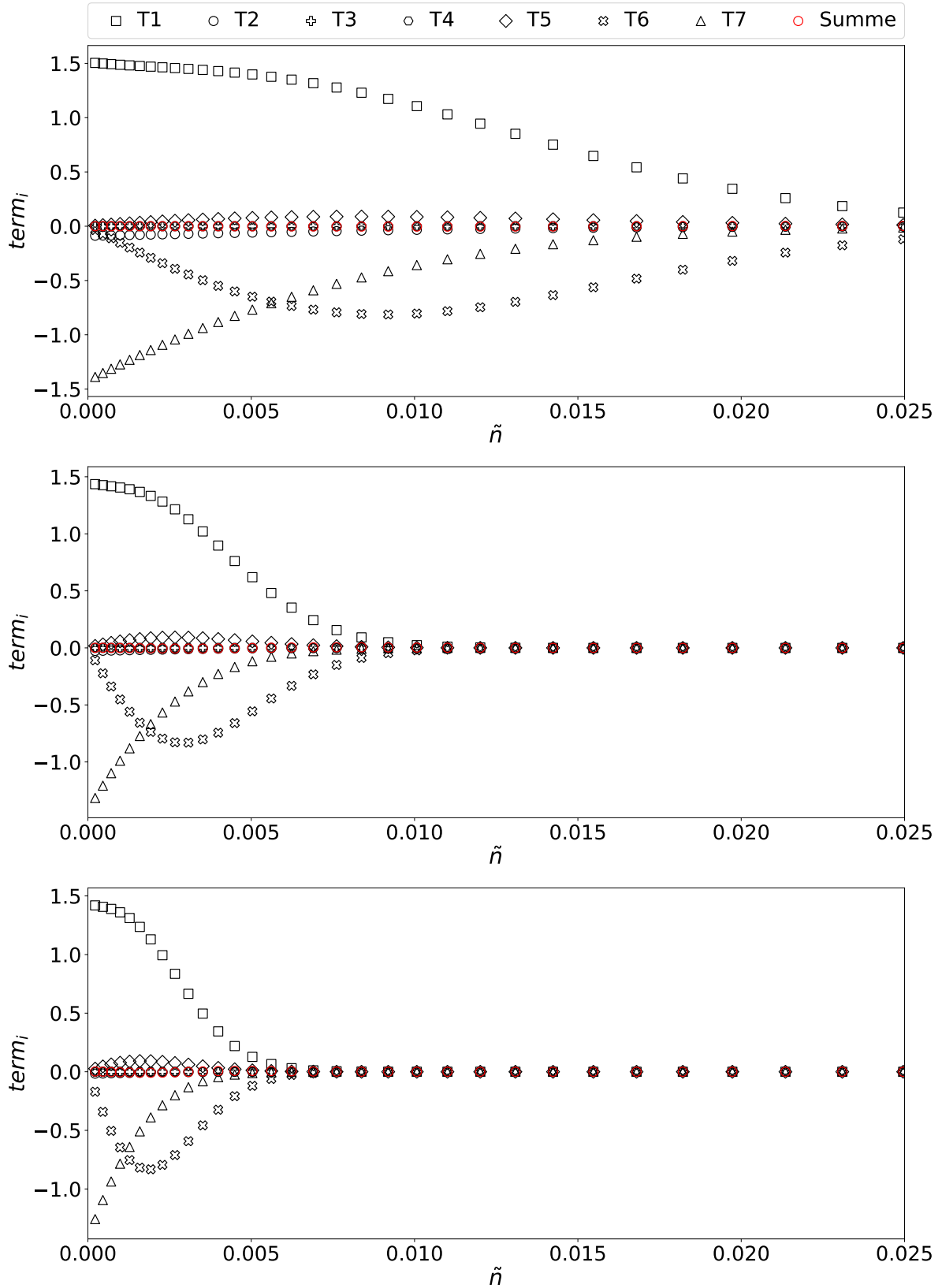


Abbildung 19: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A bei unterschiedlichen Péclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$).

Als Nächstes soll aus Gleichung (25) ein eindimensionales Berechnungsmodell für die Konzentrationsverteilung der drei Spezies in Normalenrichtung \tilde{n} abgeleitet werden. Dementsprechend müssen die Ableitungen bezüglich der Zeit \tilde{t} und

der Normalenrichtung \tilde{n} mit finiten Differenzen approximiert werden. Die globalen Parameter Damköhler-Zahl Da und Péclet-Zahl Pe sind Variationsparameter und werden jeweils vorgegeben. Modelliert werden müssen schließlich noch $\tilde{U}_{\tilde{n}}$, $\tilde{U}_{\tilde{\tau}}$ und $\frac{\partial \tilde{c}_i}{\partial \tilde{\tau}}$. Dazu werden diese Terme zuerst analysiert. Betrachtet wird zuerst die Normalgeschwindigkeit $\tilde{U}_{\tilde{n}}$. Wie in Abbildung 20 zu erkennen ist, nimmt diese linear mit dem Abstand zur Grenzfläche zu. Die Steigung $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ ist abhängig vom Polarwinkel Θ . Für $\Theta < 90^\circ$ ist diese negativ. Für $\Theta > 90^\circ$ ist sie positiv. In Abbildung 23 ist dies deutlich dargestellt.

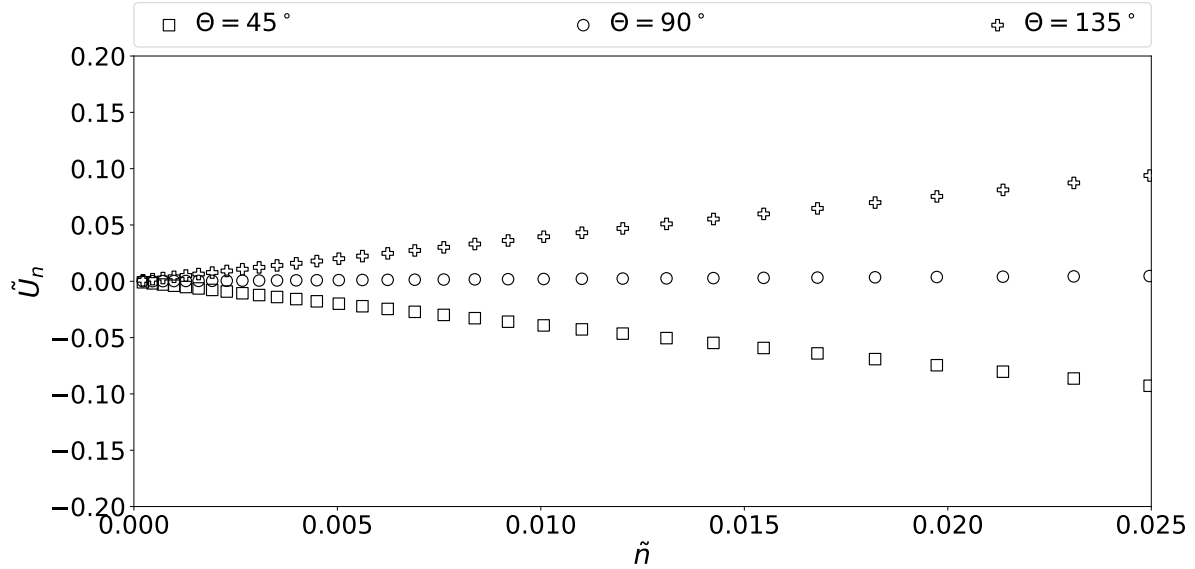


Abbildung 20: Verlauf der Geschwindigkeit $\tilde{U}_{\tilde{n}}$ in Normalenrichtung für unterschiedliche Polarwinkel Θ ($Pe = 125000$).

Als Nächstes wird die Geschwindigkeit $\tilde{U}_{\tilde{\tau}}$ betrachtet. Wie Abbildung 21 veranschaulicht, ist der Geschwindigkeitsbetrag in dem betrachteten Bereich abhängig vom Polarwinkel Θ aber nicht vom Abstand \tilde{n} zur Grenzfläche. $\tilde{U}_{\tilde{\tau}}$ folgt näherungsweise dem Verlauf einer Sinus-Funktion und lässt sich als solche modellieren.

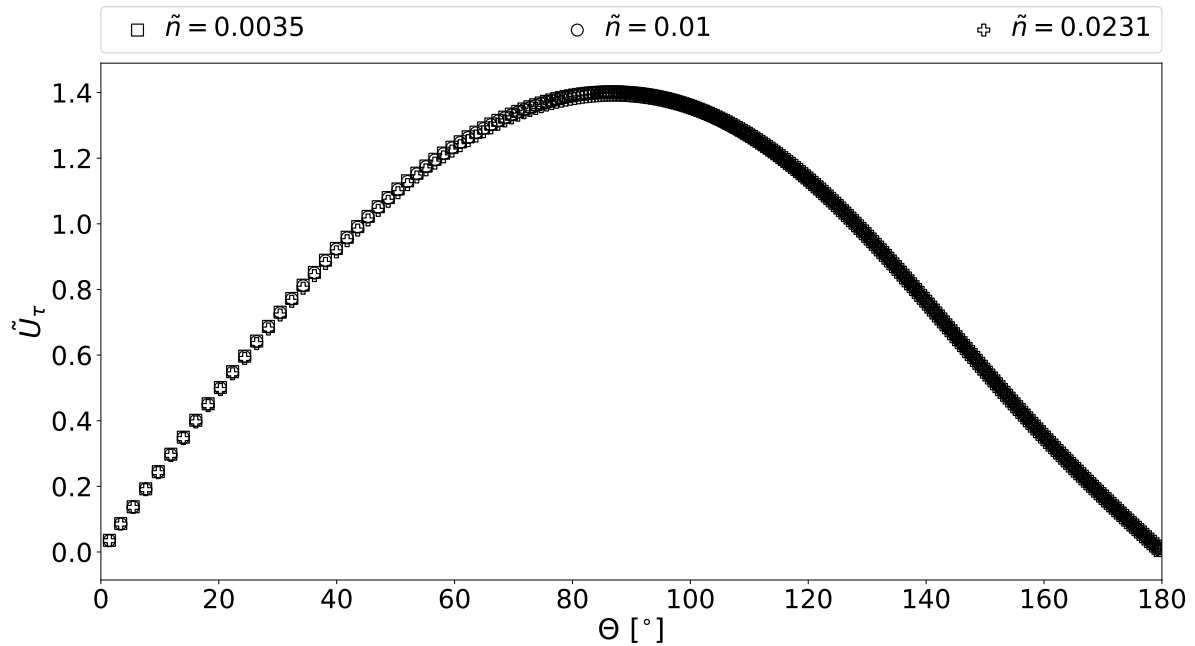


Abbildung 21: Verlauf der Tangentialgeschwindigkeit $\tilde{U}_{\tilde{\tau}}$ über dem Polarwinkel Θ für unterschiedliche Abstände \tilde{n} ($Pe = 125000$).

In Abbildung 22 sind die Verläufe des Terms $\frac{\partial \tilde{c}_A}{\partial \tilde{r}}$ in Normalenrichtung (oben) und in tangentialer Richtung (unten) dargestellt. Wie zu erkennen ist, sind diese Verläufe komplexer. Da im Rahmen dieser Arbeit kein einfacher Modellierungsansatz für $\frac{\partial \tilde{c}_i}{\partial \tilde{r}}$ gefunden wurde, wird der zweite Konvektionsterm erst einmal vernachlässigt. Dies stellt eine Fehlerquelle dar und bietet eine Möglichkeit die Trainingsdaten für das resultierende Subgridskalen-Modell zu verbessern.

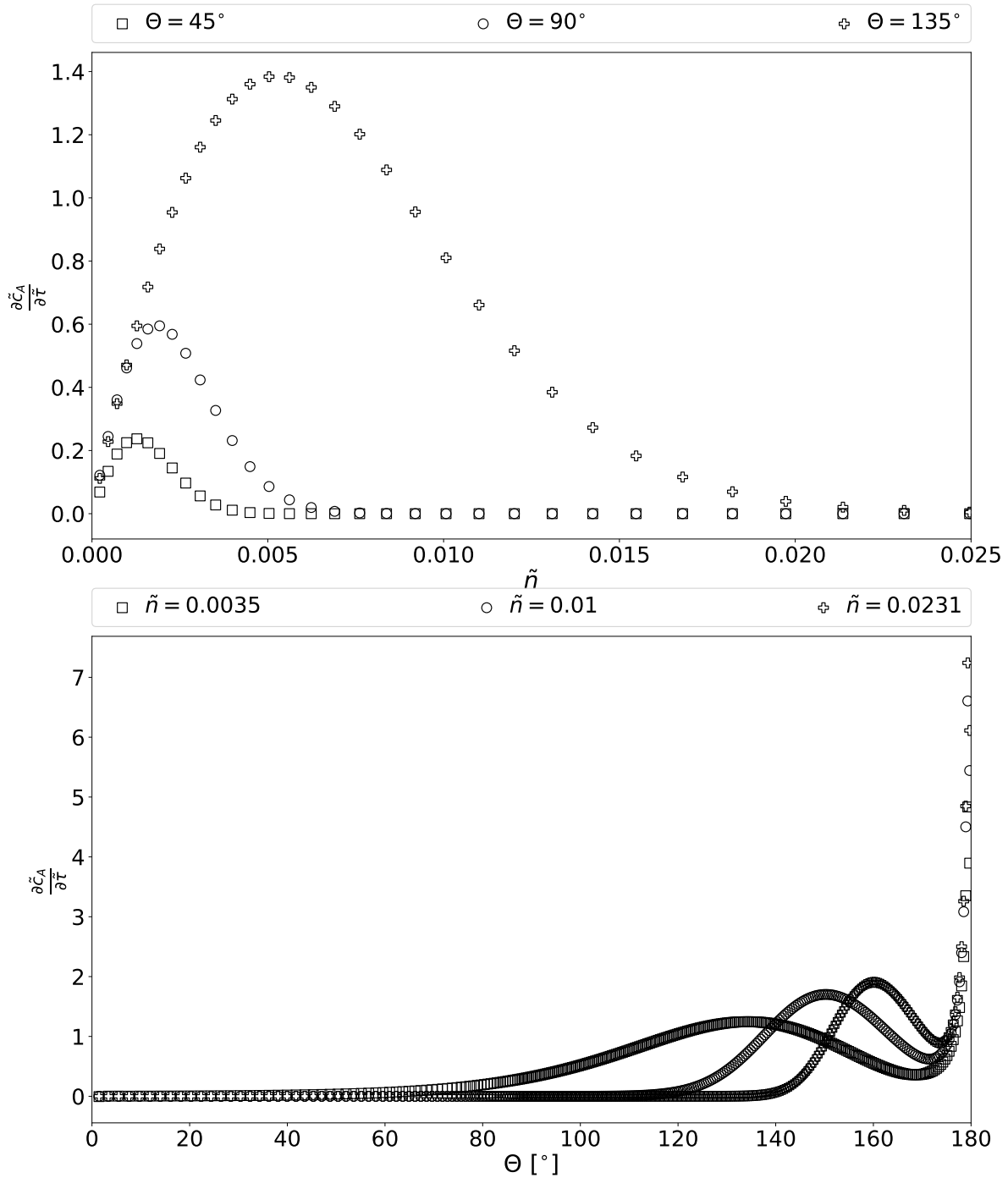


Abbildung 22: Verlauf von $\frac{\partial \tilde{c}_A}{\partial \tilde{r}}$ in Normalenrichtung für verschiedene Polarwinkel Θ (oben). Verlauf von $\frac{\partial \tilde{c}_A}{\partial \tilde{r}}$ in tangentialer Richtung für verschiedene Abstände \tilde{n} (unten). ($Pe = 125000$; $Da = 0.3$; $c_{B,0} = 5$).

Term-Modellierung

Durch Weglassen des zweiten Konvektionsterms muss nur noch $\tilde{U}_{\tilde{n}}$ modelliert werden. Wie schon erwähnt, ist die Geschwindigkeit $\tilde{U}_{\tilde{n}}$ in dem betrachteten Bereich linear und die Steigung vom Polarwinkel Θ abhängig. Da keiner der anderen Terme aus Gleichung (27) eine Abhängigkeit vom Polarwinkel aufweist, kann die Steigung $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ direkt als Variationsparameter benutzt werden. Eine direkte Abhängigkeit von Θ entfällt. $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ geht demnach als neuer Parameter in das eindimensionale Modell ein. Die Geschwindigkeit in Normalenrichtung wird also nach Gleichung (26) modelliert.

$$\tilde{U}_{\tilde{n}} = \tilde{n} \partial_{\tilde{n}} \tilde{U}_{\tilde{n}} \big|_{\tilde{n}=0} \quad (26)$$

$$\frac{\partial \tilde{c}_i}{\partial \tilde{t}} = \frac{1}{Pe} \frac{\partial^2 \tilde{c}_i}{\partial \tilde{n}^2} - \tilde{U}_{\tilde{n}} \frac{\partial \tilde{c}_i}{\partial \tilde{n}} \pm Da \tilde{c}_A \tilde{c}_B \quad (27)$$

Es muss nun noch ein Variationsbereich für den neuen Variationsparameter $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ definiert werden. Da $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ für einen bestimmten Winkel Θ konstant ist, ist es ausreichend die Werte an der Grenzfläche zu betrachten. In Abbildung 23 sind die Verläufe für $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}} \big|_{\Sigma}$ über dem Polarwinkel für verschiedene Reynolds-Zahlen dargestellt. Wie zu erkennen ist, bewegen sich die Werte im Intervall $[-6, 5]$. Dieses Intervall wird als Variationsbereich genommen.

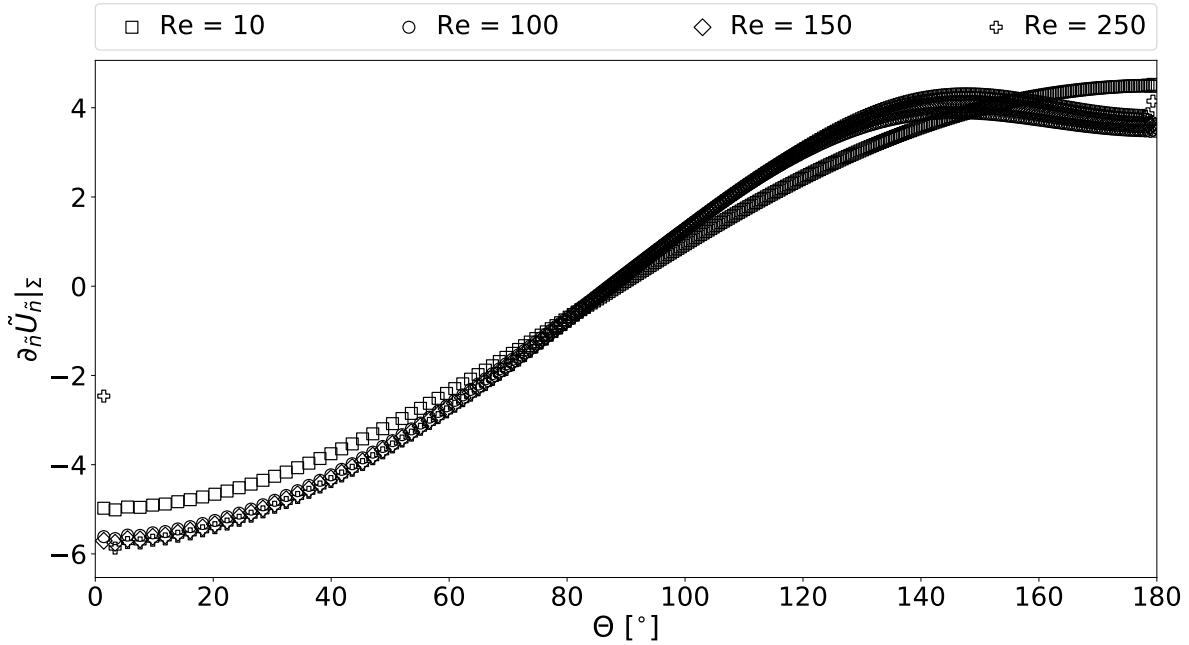


Abbildung 23: Verlauf von $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ an der Grenzfläche über dem Polarwinkel Θ für verschiedene Reynolds-Zahlen.

Numerische Lösung

Das Modell besteht schließlich aus Gleichung (27). Diese Gleichung wird mit dem Crank-Nicolson-Verfahren gelöst (Details siehe Anhang C). Mit den Vereinfachungen $f = \frac{\delta t}{Pe (\delta \tilde{n})^2}$, $g = \frac{\delta \tilde{t}}{2 \delta \tilde{n}}$ und $h = Da \delta \tilde{t}$ ergibt sich Gleichung 28 für Spezies A, wobei $\delta \tilde{n} = \tilde{L}/(N_n - 1)$ die örtliche und $\delta \tilde{t} = \tilde{T}/(N_t - 1)$ die zeitliche Auflösung beschreiben. Die Gleichungen für Spezies B und P können analog hergeleitet werden. \tilde{c}_j^i beschreibt die Konzentration von Spezies A an der Stelle j zur Zeit i . N_n ist die Anzahl an Gitterpunkten (Anmerkung: es handelt sich um ein 1D-Gitter) und N_t die Anzahl an berechneten Zeitpunkten. Die Endzeit der Berechnung wird auf $\tilde{T} = 5$ gesetzt, vgl. Kapitel 3.3. Die Größe des Berechnungsgebietes des 1D-Modells wird zu $\tilde{L} = 0.03$ gesetzt.

$$\begin{aligned} & \left(\tilde{U}_{\tilde{n},j} \frac{g}{2} - \frac{f}{2} \right) \tilde{c}_{j+1}^{i+1} + \left[1 + f + \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i) \right] \tilde{c}_j^{i+1} + \left(-\tilde{U}_{\tilde{n},j} \frac{g}{2} - \frac{f}{2} \right) \tilde{c}_{j-1}^{i+1} \\ & = \left(-\tilde{U}_{\tilde{n},j} \frac{g}{2} + \frac{f}{2} \right) \tilde{c}_{j+1}^i + \left[1 - f - \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i) \right] \tilde{c}_j^i + \left(\tilde{U}_{\tilde{n},j} \frac{g}{2} + \frac{f}{2} \right) \tilde{c}_{j-1}^i \end{aligned} \quad (28)$$

Rand	Spezies A	Spezies B	Spezies P
$\tilde{n} = 0$	$\tilde{c}_A = 1$	$\frac{\partial \tilde{c}_B}{\partial \tilde{n}} = 0$	$\frac{\partial \tilde{c}_P}{\partial \tilde{n}} = 0$
$\tilde{n} = \tilde{L}$	$\frac{\partial \tilde{c}_A}{\partial \tilde{n}} = 0$	$\frac{\partial \tilde{c}_B}{\partial \tilde{n}} = 0$	$\frac{\partial \tilde{c}_P}{\partial \tilde{n}} = 0$

Tabelle 11: Randbedingungen für eindimensionales Modell.

Diese Gleichung gilt für jeden Punkt j . Werden alle N_n Punkte und ihre jeweiligen Beziehungen zueinander betrachtet, so ergibt sich ein Gleichungssystem mit Matrizen und Vektoren, siehe Gleichung (29). Es werden die Abkürzungen aus Tabelle 25 verwendet. Hierbei ist zu beachten, dass die Terme $a_{1,j}$ bis $a_{4,j}$ immer abhängig von der Stelle j sind. In diese Gleichung sind bereits die Randbedingungen für die einzelnen Spezies eingearbeitet (siehe erste und letzte Zeile der Matrizen). Die Randbedingungen sind in Tabelle 11 zusammengefasst.

Eintrag	Term
$a_{1,j}$	$\tilde{U}_{\tilde{n},j} \frac{g}{2} - \frac{f}{2}$
$a_{2,j}$	$1 + f + \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i)$
$a_{3,j}$	$-\tilde{U}_{\tilde{n},j} \frac{g}{2} - \frac{f}{2}$
$a_{4,j}$	$1 - f - \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i)$

Tabelle 12: Abkürzungen

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ a_{3,1} & a_{2,1} & a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_{3,N_n-1} & a_{2,N_n-1} & a_{1,N_n-1} \\ 0 & \cdots & \cdots & 0 & -f & a_{2,N_n} \end{bmatrix} \mathbf{c}_A^{i+1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -a_{3,1} & a_{4,1} & -a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -a_{3,N_n-1} & a_{4,N_n-1} & -a_{1,N_n-1} \\ 0 & \cdots & \cdots & 0 & f & a_{4,N_n} \end{bmatrix} \mathbf{c}_A^i \quad (29)$$

Analog zur Gleichung für Spezies A, $A \cdot \mathbf{c}_A^{i+1} = M \cdot \mathbf{c}_A^i$ (vgl. Gleichung 29), können auch Gleichungssysteme für Spezies B und P formuliert werden, siehe Gleichung (30). Die Matrizen B , N , P und O sind im Anhang D detailliert dargestellt.

$$\begin{aligned} B \cdot \mathbf{c}_B^{i+1} &= N \cdot \mathbf{c}_B^i \\ P \cdot \mathbf{c}_P^{i+1} &= O \cdot \mathbf{c}_P^i + \mathbf{RT} \\ \mathbf{RT}_j &= \frac{h}{4} (\tilde{c}_{A,j}^{i+1} + \tilde{c}_{A,j}^i) (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i) \end{aligned} \quad (30)$$

Die Gleichungssysteme (29) und (30) sind aufgrund des Reaktionsterms in Gleichung (27) alle miteinander gekoppelt. Dies drückt sich in den Einträgen $a_{2,j}$ und $a_{4,j}$ aus (siehe Tabelle 25). Dies bedeutet, dass die Matrizen A und M aus Gleichung (29) von \mathbf{c}_B^{i+1} abhängig sind, welches noch zu berechnen ist. Analog sind die Matrizen B und N aus Gleichung (30) von \mathbf{c}_A^{i+1} abhängig. Dieses Kopplungsproblem wird iterativ gelöst. Für die Berechnung jedes Zeitschritts $i+1$ werden im ersten Iterationsschritt sowohl $\mathbf{c}_A^{i+1*} = \mathbf{c}_A^i$ als auch $\mathbf{c}_B^{i+1*} = \mathbf{c}_B^i$ gesetzt. Mit dieser Annahme wird ein neues \mathbf{c}_A^{i+1**} und \mathbf{c}_B^{i+1**} berechnet. Ist die Abweichung zwischen berechnetem und angenommenem Konzentrationsverlauf größer als eine geforderte Toleranz wird die Iteration fortgesetzt. Dieses mal mit $\mathbf{c}_A^{i+1*} = \mathbf{c}_A^{i+1**}$ und $\mathbf{c}_B^{i+1*} = \mathbf{c}_B^{i+1**}$.

Ist die iterative Berechnung von \mathbf{c}_A^{i+1} und \mathbf{c}_B^{i+1} abgeschlossen, wird die Gleichung für Spezies P gelöst. Die Kopplung dieser Gleichung mit denen für Spezies A und B erfolgt über den Term \mathbf{RT} . Für diesen gilt der Zusammenhang aus Gleichung (30). Da diese Werte schon vorliegen, kann \mathbf{c}_P^{i+1} ohne Iteration berechnet werden. In Algorithmus 1 ist Lösung des gekoppelten Problems für einen Zeitschritt $i+1$ als Pseudo-Code dargestellt.

Algorithm 1 Kopplung Spezies Berechnung

```
1: procedure BERECHNUNG ZEITSCHRITT  $i + 1(\mathbf{c}_A^i, \mathbf{c}_B^i, \mathbf{c}_P^i)$ 
2:   select  $\mathbf{c}_A^{i+1*} = \mathbf{c}_A^i$  und  $\mathbf{c}_B^{i+1*} = \mathbf{c}_B^i$ 
3:   set  $diff > tol$ 
4:   while  $diff > tol$  do
5:     calculate  $[A^*, M^*, B^*, N^*] = F(\mathbf{c}_A^{i+1*}, \mathbf{c}_A^i, \mathbf{c}_B^{i+1*}, \mathbf{c}_B^i)$ 
6:     solve  $A^* \cdot \mathbf{c}_A^{i+1**} = M^* \cdot \mathbf{c}_A^i$ 
7:     solve  $B^* \cdot \mathbf{c}_B^{i+1**} = N^* \cdot \mathbf{c}_B^i$ 
8:     calculate  $diff = f(\mathbf{c}_A^{i+1**}, \mathbf{c}_A^{i+1*}, \mathbf{c}_B^{i+1**}, \mathbf{c}_B^{i+1*})$ 
9:     update  $\mathbf{c}_A^{i+1*} = \mathbf{c}_A^{i+1**}$ 
10:    update  $\mathbf{c}_B^{i+1*} = \mathbf{c}_B^{i+1**}$ 
11:  set  $\mathbf{c}_A^{i+1} = \mathbf{c}_A^{i+1**}$ 
12:  set  $\mathbf{c}_B^{i+1} = \mathbf{c}_B^{i+1**}$ 
13:  calculate  $P, O$  und  $RT = F(\mathbf{c}_A^{i+1}, \mathbf{c}_A^i, \mathbf{c}_B^{i+1}, \mathbf{c}_B^i)$ 
14:  solve  $P \cdot \mathbf{c}_P^{i+1} = O \cdot \mathbf{c}_P^i + RT$ 
15:  return  $\mathbf{c}_A^{i+1}, \mathbf{c}_B^{i+1}, \mathbf{c}_P^{i+1}$ 
```

Schließlich wird für die Lösung dieses Gleichungssystems noch eine Anfangsbedingung benötigt (siehe Tabelle 13). Zu Beginn liegt Spezies A nur an der Grenzfläche vor, Spezies B nur im Fluid. Spezies P ist zu Beginn der Rechnung noch nicht vorhanden.

Rand	Spezies A	Spezies B	Spezies P
$\tilde{n} = 0$	$\tilde{c}_A = 1$	$\tilde{c}_B = 0$	$\tilde{c}_P = 0$
$\tilde{n} > 0$	$\tilde{c}_A = 0$	$\tilde{c}_B = c_{B0}$	$\tilde{c}_P = 0$

Tabelle 13: Anfangsbedingungen für eindimensionales Modell.

Aus den bisher berechneten Konzentrationsverläufen für die drei Spezies für unterschiedliche Zeitschritte müssen nun noch die nötigen Features und Label berechnet werden. Die Herleitung und Bedeutung der Label ist in Abschnitt 2.4 erklärt. Die Features sind die Variationsparameter Pe , Da , CB , $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ und der Abstand \tilde{n} . Außerdem gehören zu den Features die durchschnittlichen Konzentrationen $\tilde{c}_i|_{\tilde{l}}$ von der Grenzfläche bis zu einem bestimmten Abstand \tilde{l} von der Grenzfläche für alle betrachteten Abstände \tilde{n} . Da die Gitterpunkte äquidistant verteilt sind, kann für einen Abstand \tilde{l} die durchschnittliche Konzentration einfach als arithmetischer Mittelwert aller Konzentrationen mit Abstand $\tilde{n} \leq \tilde{l}$ berechnet werden.

Die Konzentrationsgradienten in Normalenrichtung werden als finite Differenzen berechnet. So wird die Normalenableitung an der Grenzfläche $\partial_{\tilde{n}} \tilde{c}_i|_{\Sigma}$ berechnet als $(\tilde{c}_1 - \tilde{c}_0)/\delta\tilde{n}$. Normalenableitungen $\partial_{\tilde{n}} \tilde{c}_i|_{\tilde{l}}$ mit Abstand \tilde{l} zur Grenzfläche werden nach $(\tilde{c}_{j+1} - \tilde{c}_{j-1})/(2\delta\tilde{n})$ approximiert.

Das Label Reaktionsterm $R|_{\tilde{l}}$ für den Abstand \tilde{l} ergibt sich als Mittelwert des Produkt der Konzentrationen der Spezies A und bei B bei Abstand \tilde{l} . Die Damköhler-Zahl Da geht hier nicht ein, siehe dazu Kapitel 2.4, insbesondere Tabelle 3.

Aus diesen Größen werden schließlich die Datenpunkte erzeugt. Ein Datenpunkt ist ein Vektor, dessen Einträge die Features und Label sind. In Abschnitt 4 Gleichung (32) ist beispielhaft ein Datenpunkt dargestellt.

Studie: Auflösung Ort/Zeit für 1D-Modell

In das Crank-Nicolson-Verfahren gehen auch die örtliche $\delta \tilde{n}$ und zeitliche Auflösung $\delta \tilde{t}$ ein. Diese werden über die Anzahl an Gitterpunkten N_n und Zeitpunkten N_t festgelegt, welche es zu bestimmen gilt. Bei dem Crank-Nicolson-Verfahren kann es zu Stabilitätsproblemen kommen und Oszillationen auftreten [7]. Dies wird deutlich in Abbildung 24. Außerdem ist zu erkennen, dass Oszillationen für gleiches N_n und N_t bei kleineren Péclet-Zahlen Pe auftreten und stärker werden. Die Relation von $\delta \tilde{t}$ zu $\delta \tilde{n}$ spielt ebenfalls eine Rolle (Diffusionszahl).

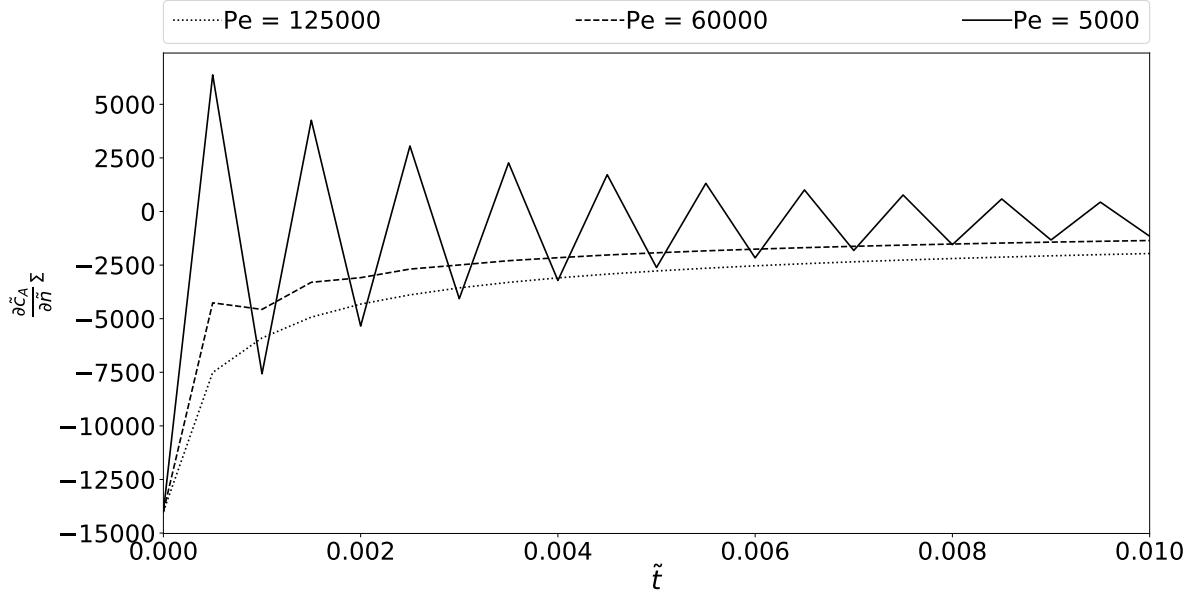


Abbildung 24: Auftretende Oszillationen in der Crank-Nicolson-Lösung, bei $N_n = 421$ und $N_t = 10001$.

Es werden mehrere Berechnungen mit verschiedenen $N_n \in [271, 301, 331, 361, 391, 421]$ und $N_t \in [5001, 10001, 15001, 25001, 50001, 75001]$ durchgeführt. Dies geschieht jeweils für die Parameterkombinationen $\{Pe = 5000, Da = 0, \tilde{c}_{B,0} = 5, \partial_{\tilde{n}} \tilde{U}_{\tilde{n}} = 5\}$, $\{Pe = 60000, Da = 0.15, \tilde{c}_{B,0} = 5, \partial_{\tilde{n}} \tilde{U}_{\tilde{n}} = 5\}$ und $\{Pe = 125000, Da = 0.3, \tilde{c}_{B,0} = 5, \partial_{\tilde{n}} \tilde{U}_{\tilde{n}} = 5\}$. Diese decken den in Kapitel 3.3 gewählten Parameterbereich ab. Betrachtet wird der Konzentrationsgradient an der Grenzfläche $\frac{\partial \tilde{c}_A}{\partial \tilde{n}} \big|_{\Sigma}$.

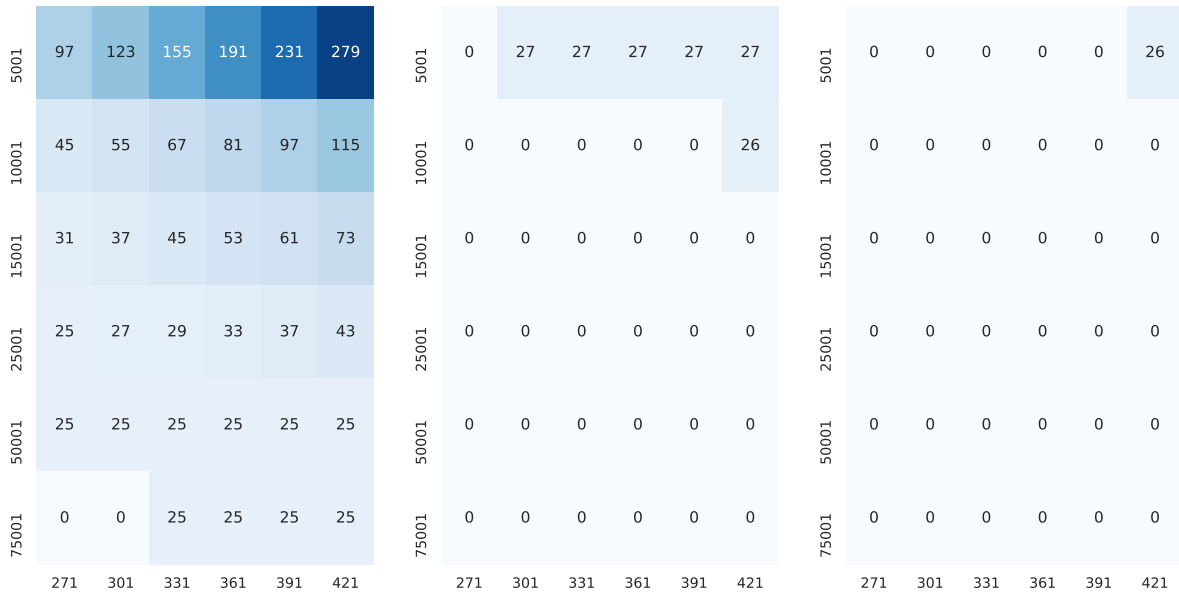


Abbildung 25: Auftreten von Oszillationen für $Pe = 5000$ (links), $Pe = 60000$ (Mitte), $Pe = 125000$ (rechts). Auf der x-Achse sind N_n und auf der y-Achse die N_t aufgetragen. Die Zahlen verdeutlichen bis zu welchem Zeitschritt Oszillationen auftreten.

In Abbildung 25 sind die Ergebnisse der unterschiedlichen N_n und N_t Kombinationen festgehalten. Die Zahlen geben den Zeitschritt an, bei dem die Amplitude im Verhältnis zum aktuellen Mittelwert kleiner als 1% ist. Bei einem Eintrag 0 liegt ein monotoner Verlauf wie in Abbildung 24 für $Pe = 125000$ vor. Hier ist noch deutlicher zu erkennen, dass bei kleineren Pe die Oszillationen zunehmen. Dies kann auch, wie schon in Abbildung 24 zu erkennen ist, zu unphysikalischen Verläufen, wie $\frac{\partial \zeta_A}{\partial \tilde{t}} \Big|_{\Sigma} > 0$, führen. Berechnet man den Mittelwert zwischen zwei Zeitschritten einer oszillierenden Lösung führt dies zu deutlich realistischeren Verläufen. Siehe dazu Abbildung 26.

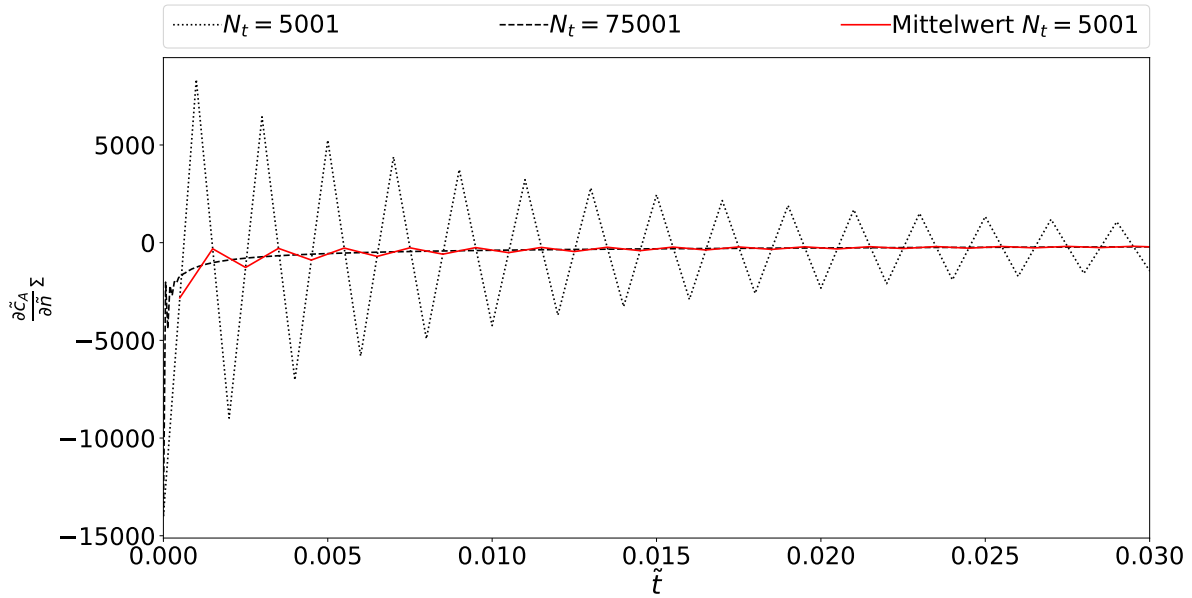


Abbildung 26: Vergleich zwischen Mittelwert-Verlauf und oszillierendem Verlauf bei $Pe = 5000$ und $N_n = 421$.

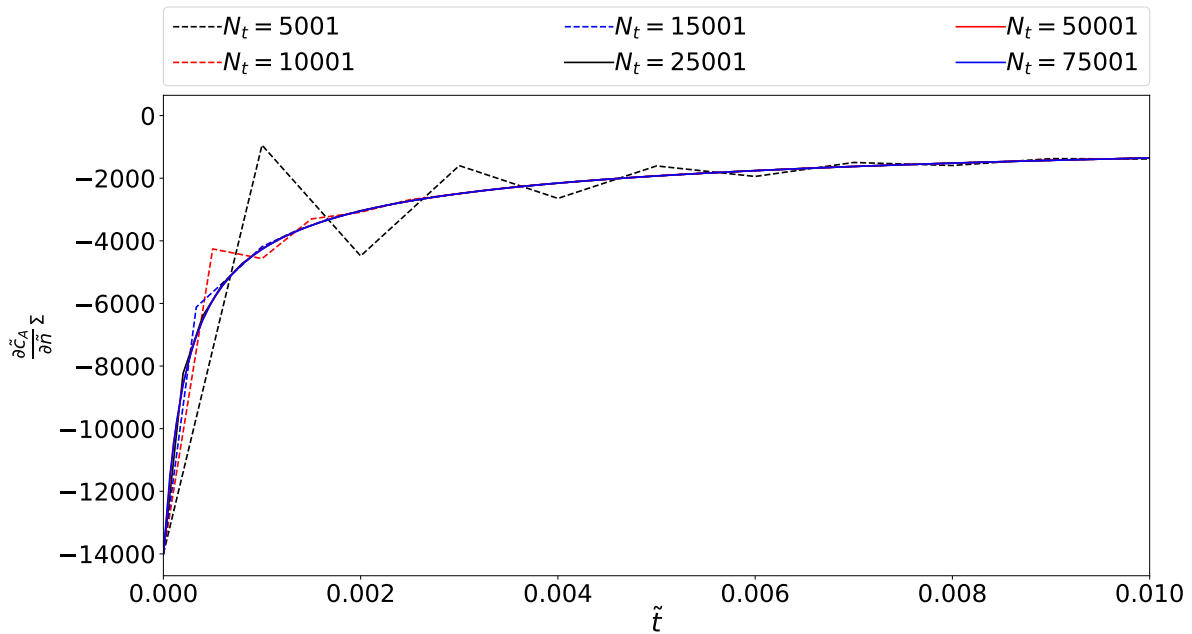


Abbildung 27: Einfluss der Zeitschritt-Anzahl bei gleichbleibendem N_n auf die Oszillation. $Pe = 60000$, $Da = 0.15$, $N_n = 421$

Um die Berechnungszeit möglichst gering zu halten, sollen N_n und N_t möglichst klein gewählt werden, jedoch groß genug um Oszillationen zu verhindern bzw. klein zu halten. Anhand der Abbildung 25 wird $N_t = 10001$ gewählt. Um für $Pe = 5000$ Oszillationen zu verhindern, müssen sehr große N_t gewählt werden, wodurch die Datenmenge und Berechnungszeit

stark ansteigt. Für kleinere Péclet-Zahlen wird dementsprechend der Ansatz verfolgt aus den berechneten Werten für die Konzentration die Mittelwerte zu berechnen und diese zu verwenden. Für $Pe = 125000$ und $Pe = 60000$ treten kaum Oszillationen auf.

Abbildung 27 verdeutlicht, wie mit größerem N_t die Oszillationen zurückgehen und sich die Verläufe immer mehr der Lösung $N_n = 421$ und $N_t = 75001$ annähern. Außerdem ist zu erkennen, dass für fortschreitende Zeit \tilde{t} die Oszillationen abklingen und sich dem gleichen Verlauf annähern. Die Wahl von N_t hat somit hauptsächlich einen Einfluss auf die Genauigkeit der Werte bei kleinen Zeiten \tilde{t} .

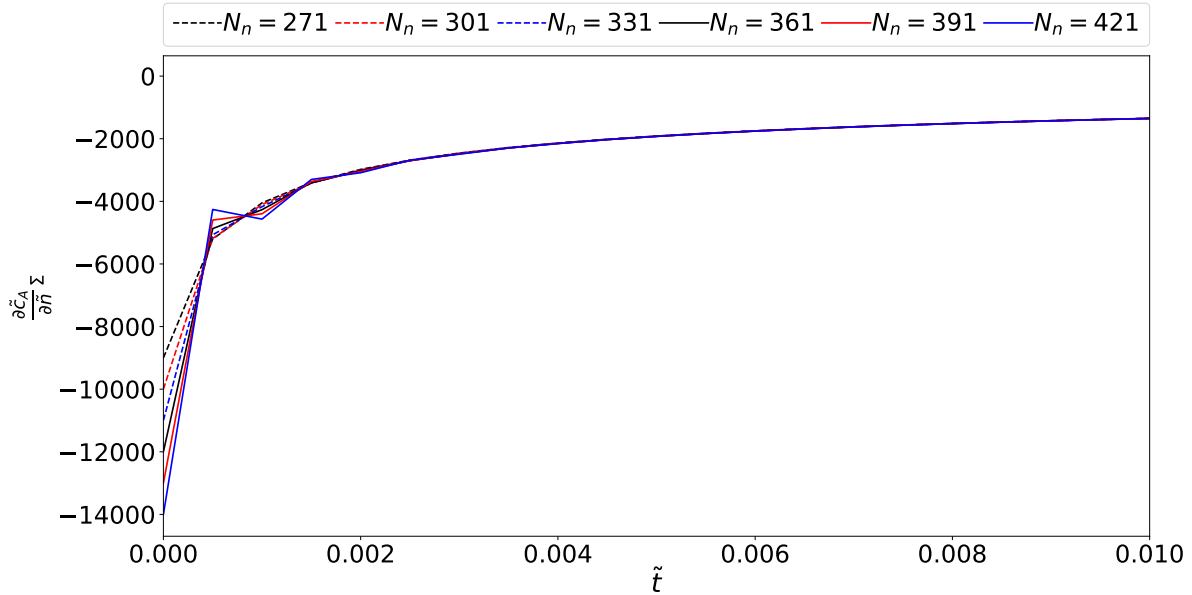


Abbildung 28: Einfluss der Gitterpunkt-Anzahl bei gleichbleibendem N_t . $Pe = 60000, Da = 0.15, N_n = 10001$

Abbildung 28 zeigt deutlich, dass auch eine Abhängigkeit von N_n besteht. Für $\tilde{t} = 0$ wird $\frac{\partial c_A}{\partial n} \big|_{\Sigma}$ erwartungsgemäß immer größer mit mehr Gitterpunkten. Die Anfangsbedingung für die Berechnung entspricht einem Konzentrationssprung. Somit werden nur Punkte mit $\tilde{t} > 0$ betrachtet. Die Berechnungen mit $N_n = 421$ und $N_t = 75001$ haben die genauesten Verläufe und werden dementsprechend als Referenzen verwendet. N_n wird anhand eines Vergleiches mit diesen bestimmt. Dafür wird der durchschnittliche Fehler F_D bis zur Zeit $\tilde{t} = 0.01$ berechnet. Dieser berechnet sich nach Gleichung (31), wobei N_i die Anzahl der betrachteten Zeitpunkte \tilde{t}_i ist und mit int die Interpolation des Referenz-Verlaufs ($N_n = 421, N_t = 75001$) bezeichnet wird. Bei $Pe = 5000$ wurden die Mittelwerte verwendet, da hier bei allen Kombinationen Oszillationen vorliegen. Für $Pe = 125000$ ergibt sich der Zusammenhang, dass mit größer werdendem N_n der durchschnittliche Fehler abnimmt. Für $Pe = 5000$ ergibt sich der genau gegenteilige Zusammenhang. Bei $Pe = 60000$ lässt sich der kleinste durchschnittliche Fehler bei $N_n = 331$ ermitteln (siehe Tabelle 14). Dieser Wert wird genommen.

$$F_D = \frac{1}{N_i} \sum_i^{N_i} \left| \frac{\partial c_A}{\partial n} \big|_{\Sigma}(\tilde{t}_i, N_n, N_t = 10001) - \left(\frac{\partial c_A}{\partial n} \big|_{\Sigma} \right)^{int}(\tilde{t}_i, N_n = 421, N_t = 75001) \right| \quad (31)$$

N_n	$Pe = 5000$	$Pe = 60000$	$Pe = 125000$
271	136,31	67,48	145,80
301	159,76	61,42	108,07
331	182,30	60,80	81,14
361	203,80	64,32	62,43
391	224,21	83,65	50,15
421	243,48	153,81	42,98

Tabelle 14: Durchschnittliche Fehler F_D für $N_t = 10001$ im Vergleich zu $N_n = 421, N_t = 75001$.

Für die Berechnung der Trainingsdaten wird $N_n = 331$ und $N_t = 10001$ verwendet.

4 Datenreduktion

4.1 Übersicht

Nach der Datengenerierung besteht der Datensatz nun aus vielen Datenpunkten, deren Komponenten mögliche Features und Label sind. So sind die ersten vier Einträge die Parameter, die für die Eingabe in das 1D-Modell benötigt werden, Pe , Da , $\tilde{c}_{B,0}$ und $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$. Weitere Einträge sind der Abstand \tilde{n} zur Grenzfläche Σ und die, von der Grenzfläche Σ bis zu diesem Abstand, durchschnittliche Konzentration der jeweiligen Spezies \tilde{c}_i . Die Label-Einträge sind die Normalableitung der Konzentration an der Grenzfläche für Spezies A $(\partial_{\tilde{n}}\tilde{c}_A|_{\Sigma})$ und für alle drei Spezies die Normalableitungen bei einem Abstand \tilde{l} $(\partial_{\tilde{n}}\tilde{c}_i|_{\tilde{l}})$. Für die Korrekturen der advektiven Flüsse werden die Konzentrationen $(\tilde{c}_i|_{\tilde{l}})$ der drei Spezies bei einem Abstand \tilde{l} benötigt. Außerdem wird noch der Wert des Reaktionsterm-Label $(R|_{\tilde{l}} = \tilde{c}_A\tilde{c}_B|_{\tilde{l}})$ bei Abstand \tilde{l} gebraucht. In Gleichung (32) ist beispielhaft ein Datenpunkt für einen beliebigen Abstand \tilde{l} dargestellt.

$$(Pe, Da, \tilde{c}_{B,0}, \partial_{\tilde{n}}\tilde{U}_{\tilde{n}}, \tilde{l}, \tilde{c}_A|_{\tilde{l}}, \tilde{c}_B|_{\tilde{l}}, \tilde{c}_P|_{\tilde{l}}, \partial_{\tilde{n}}\tilde{c}_A|_{\Sigma}, \partial_{\tilde{n}}\tilde{c}_A|_{\tilde{l}}, \tilde{c}_A|_{\tilde{l}}, \partial_{\tilde{n}}\tilde{c}_B|_{\tilde{l}}, \tilde{c}_B|_{\tilde{l}}, \partial_{\tilde{n}}\tilde{c}_P|_{\tilde{l}}, \tilde{c}_P|_{\tilde{l}}, R|_{\tilde{l}}) \quad (32)$$

Die Datenreduktion soll mehrfach vorkommende Datenpunkte auf einen repräsentativen Datenpunkt reduzieren. Dies schließt auch Punkte ein, deren Komponenteneinträge nicht exakt übereinstimmen, die aber sehr nahe beieinander liegen. Solche Anhäufungen von Datenpunkten entstehen zum Beispiel durch das Weglassen der Zeit als Feature. Nach Erreichen eines stationären Zustandes bei der Modellrechnung werden mit jedem Zeitschritt Datenpunkte reproduziert. Außerdem können nach der späteren Feature-Auswahl durch Weglassen von Features und der damit verbundenen Projektion der Datenpunkte in einen niedriger dimensional Raum Anhäufungen von Datenpunkten entstehen. Diese können durch eine nochmalige Reduktion ebenfalls herausgefiltert werden. Eine Übersicht aller Arbeitsschritte von der Datengenerierung bis hin zum fertigen SGS-Modell folgt in Kapitel 5.1.

Solche Anhäufungen von Datenpunkten können das Ergebnis der folgenden Regression beeinflussen. Beim Training des neuronalen Netzes werden anhand einer Verlustfunktion die Datenpunkte approximiert. Die Minimierung der Verlustfunktion kann nun, aufgrund der Menge an Datenpunkten, zu einer genaueren Approximation im Bereich dieser Anhäufungen führen. Gleichzeitig kann es zu einer Verschlechterung der Vorhersagegenauigkeit in Bereichen mit weniger Datenpunkten kommen. Einen ähnlichen Effekt haben Anhäufungen auch auf die statistischen Verfahren für die Analyse der Features (Korrelationen zwischen Features). Ziel der Datenreduktion ist es eine ausgeglichene Verteilung der Datenpunkte zu erreichen und größere Ansammlungen zu vermeiden. Außerdem kann so der Datensatz mit geringem Informationsverlust erheblich reduziert werden. Dies wirkt sich positiv auf den Speicherbedarf und die Rechenzeit für die nachfolgenden Schritte aus.

Zuerst noch eine kurze Übersicht über den Ablauf einer Reduktion bevor in den folgenden Unterkapiteln zwei Ideen, wie Datenreduktionsalgorithmen aussehen können, vorgestellt werden. Anschließend werden diese an einem zufalls-generierten Datensatz (2D) getestet und miteinander verglichen. Zum Schluss werden die zwei Verfahren noch an Daten des 1D-Modells getestet, dabei werden Reduktionsraten verglichen und eine einfache Abschätzung für die Dauer der jeweiligen Algorithmen präsentiert.

Ablauf

Die Reduktion der Daten des 1D-Modells wird in mehreren Schritten vollzogen. Vor der Reduktion werden die Datenpunkte skaliert. Dies geschieht für jedes Feature/Label einzeln. Die Einträge werden mit dem Minimum und Maximum des jeweiligen Features/Label skaliert, sodass alle Einträge zwischen 0 und 1 liegen. In einem ersten Reduktionsschritt wird für jede berechnete Kombination $(Pe, Da, \tilde{c}_{B,0}, \partial_{\tilde{n}}\tilde{U}_{\tilde{n}})$ eine Reduktion durchgeführt. Dies geschieht für jeden Abstand \tilde{l} individuell, da die Datenpunkte durch die Berechnungsweise schon getrennt für jeden Abstand vorliegen. Anschließend werden alle verbliebenen Datenpunkte zu einem Datensatz zusammengefügt. Der Grund hierfür ist, dass die Datenpunkte nach der Erzeugung in kleinen Datensätzen vorliegen (z.B. für $N_t = 10001$ hat der Datensatz für jede Kombination $(Pe, Da, \tilde{c}_{B,0}, \partial_{\tilde{n}}\tilde{U}_{\tilde{n}}, \tilde{l})$ eine Größe von 10000 Datenpunkten). Diese lassen sich schneller reduzieren als ein sehr großer Datensatz. Nach der Feature-Auswahl und dem Wegfallen von Features, wird noch einmal eine Reduktion vorgenommen. Auf die Feature-Auswahl wird in Kapitel 5.2.2 eingegangen. Eine detaillierte Einordnung der Datenreduktion in den Gesamtprozess ist in Kapitel 5.1 zu finden.

4.2 Round-Verfahren

Das erste Verfahren wird im Folgenden Round-Verfahren genannt. Es basiert darauf, dass die Datenpunkte, deren Einträge auch kontinuierlich verteilt sein können, auf bestimmte Werte projiziert werden. Dies geschieht, indem alle Einträge

eines Datenpunktes auf eine bestimmte, vorzugebende Nachkommastelle gerundet werden. Dieser Schritt wird in einem Arbeitsdatensatz (*workpoints*) vollzogen, der von den Einträgen her mit dem ursprünglichen identisch ist, jedoch um einen weiteren Eintrag ergänzt wurde, der einen Rückschluss auf den ursprünglichen Datenpunkt zulässt (*Index*).

Algorithm 2 Round

```

1: procedure REDUKTION-ROUND(Datenpunkte, Dezimalstelle)
2:   Datenpunkte  $\rightarrow$  workpoints
3:   for all workpoints do
4:     round Datenpunkteinträge  $\leftarrow$  Dezimalstelle
5:   save Index als zusätzlicher Eintrag  $\rightarrow$  workpoints
6:   sort by Eintrag workpoints
7:   for all workpoints do
8:     calculate Differenz  $\leftarrow$  workpoint i, workpoint i+1
9:     if Differenz = 0 then
10:      save Index-Eintrag von workpoint i+1  $\rightarrow$  Indexliste
11:   delete multiple points Datenpunkte, Indexliste
12:   return verbliebene Punkte

```

Anschließend werden die Datenpunkte sortiert und zwar aufsteigend zuerst nach dem ersten Eintrag, dann nach dem zweiten usw. Dies geschieht für alle Einträge außer dem Index-Eintrag. Als nächstes wird jeder Datenpunkt i (DP^i) mit dem darauffolgenden Datenpunkt $i + 1$ verglichen. Es wird eine Differenz gebildet. Diese ist die Summe der betragsmäßigen Differenzen der einzelnen Einträge j ohne den Index-Eintrag, siehe Gleichung (33). Ist die Differenz zwischen beiden Punkten Null, wird der Index-Eintrag des Datenpunkts $i + 1$ in der Indexliste gespeichert. Am Ende wird mit Hilfe der Indexliste auf die ursprünglichen Datenpunkte zurückgeschlossen und diese aus dem ursprünglichen Datensatz entfernt. Durch diese Vorgehensweise geht keine Information durch das zwischenzeitliche Runden des Arbeitsdatensatzes verloren.

$$Differenz_i = \sum_j |DP_j^{i+1} - DP_j^i| \quad (33)$$

Bei dieser Ausführung wird immer der erste von gleich gerundeten Datenpunkten behalten. Durch ein anderes Wahlverfahren zwischen diesen gleichen Datenpunkten, könnte eine gleichmäßigere Verteilung der übrig bleibenden Datenpunkte erreicht werden. Ein alternatives Wahlverfahren könnte so aussehen: Man nimmt jeweils die Menge an Datenpunkten, die auf einen gleichen Punkt gerundet (projiziert) werden und wählt den ursprünglichen Datenpunkt, der den kleinsten Unterschied zum gerundeten Punkt aufweist.

Zur besseren Veranschaulichung ist dieses Verfahren in Algorithmus 2 als Pseudo-Code dargestellt. Im Anhang E ist der Ablauf des Verfahrens grafisch an einer beispielhaften Menge an Datenpunkten dargestellt.

Nachteil dieses Verfahrens ist, dass auch nach der Reduktion vereinzelt noch nahe beieinanderliegende Datenpunkte vorkommen können (solche Stellen sind z.B. in Abbildung 29 Mitte zu erkennen). Dies kann an den Grenzen zwischen Auf- und Abrunden auftreten. Außerdem ist die Einstellungsmöglichkeit für die Abstände zwischen Datenpunkten limitiert. Es kann immer nur auf Nachkommastellen gerundet werden.

4.3 Metrik-Verfahren

Das zweite Verfahren basiert auf der Abstandsberechnung zwischen zwei Datenpunkten, deshalb wird dieses Verfahren im Weiteren als Metrik-Verfahren bezeichnet. Der Ablauf des Algorithmus ist wie folgt: Ein Datenpunkt wird mit allen Bezugspunkten verglichen. Über die Metrik wird der Abstand zu jedem Bezugspunkt berechnet. Ist dieser kleiner als ein vorzugebender Schwellenwert, wird der Datenpunkt aussortiert. Ist der Abstand größer als der Schwellenwert, wird der Datenpunkt der Menge der Bezugspunkte zugeordnet. Der Algorithmus startet, indem der erste Datenpunkt der leeren Menge an Bezugspunkten zugeordnet wird. Sind alle Datenpunkte entweder den Bezugspunkten zugeordnet oder aussortiert worden, endet der Algorithmus. Die Bezugspunkte sind die Datenpunkte, welche nach der Reduktion übrig bleiben. In Algorithmus 3 ist dieser Prozess als Pseudo-Code dargestellt.

$$Abstand = \sqrt{\sum_i (g_i \Delta v_i)^2} \quad (34)$$

In Gleichung (34) ist die Berechnungsvorschrift der Metrik angegeben. Δv_i ist dabei die Differenz zweier Datenpunkte bezüglich des Eintrages i im Datenvektor. Diese wird mit einem Gewicht g_i beaufschlagt. Die Gewichte sind für jeden Eintrag einzeln frei wählbar. So besteht die Möglichkeit die Dichte der Verteilung bezüglich bestimmter Features zu variieren. Große Gewichte sorgen schon bei kleinen Differenzen in den Einträgen der Datenpunkte für große Abstände. Bei sehr kleinen Gewichten werden selbst bei großen Differenzen die Datenpunkte als sehr nahe beieinander empfunden. Außerdem ist es möglich durch große Gewichte für diskret verteilte Features, wie Pe , Da , $\tilde{c}_{B,0}$, $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ oder \tilde{n} , dafür zu sorgen, dass diese Datensätze getrennt bleiben, indem der berechnete Abstand größer als der Schwellenwert wird.

Algorithm 3 Metrik

```

1: procedure REDUKTION-METRIK(Datenpunkte,Gewichte,Schwelle)
2:   1. Datenpunkt  $\rightarrow$  Bezugspunkte
3:   for all Datenpunkte do
4:     berechne Abstand  $\leftarrow$  Bezugspunkte,Gewichte
5:     if Abstand < Schwelle then
6:       Datenpunkt aussortieren
7:     else
8:       Datenpunkt  $\rightarrow$  Bezugspunkte
9:   return Bezugspunkte

```

Ein Vorteil dieses Algorithmus ist die frei einstellbare Schwelle. Durch diese lässt sich der Grad der Reduktion einstellen, also wie viele Datenpunkte nach der Reduktion übrig sind. Dies beeinflusst außerdem den Zeitaufwand der Reduktion. Ein weiterer Vorteil sind die frei einstellbaren Gewichte. Damit ist es möglich die Reduktion bestimmter Einträge zu beeinflussen. Im Vergleich zum Round-Verfahren kommen keine benachbarten Punkte vor, wodurch eine gleichmäßigere Verteilung erreicht wird. Das Metrik-Verfahren hat den Nachteil, dass es immer langsamer wird je größer die Datenmenge und die Anzahl an Bezugspunkten ist, wobei die Anzahl an Bezugspunkten mit der gewählten Schwelle zusammenhängt. Außerdem ist es deutlich langsamer als das Round-Verfahren.

Die Auswirkungen der Variation des Schwellenwertes ist in Tabelle 15 und in Abbildung 29 zu erkennen. Dieser Wert ist frei wählbar. Durch Testen mit verschiedenen Werten kann bestimmt werden, wie viele Datenpunkte jeweils nach der Reduktion übrig bleiben und es kann eine Wahl getroffen werden. Betrachtet man das Beispiel aus Tabelle 15, erkennt man, dass in diesem Fall ein Schwellenwert von 0.01 zu niedrig gewählt ist, da nur eine geringfügige Reduktion an Datenpunkten zustande kommt.

4.4 Vergleich

Als Nächstes werden die beiden Verfahren miteinander verglichen. Dies geschieht zunächst an zweidimensionalen Zufallsdaten, um diese auch grafisch darstellen zu können. In Abbildung 29 sind oben (schwarz) die Rohdaten, in Form der 300 Zufallspunkte dargestellt. In der Mitte (rot) ist das Ergebnis nach der Reduktion mit dem Verfahren Round und dem Rundungskriterium, eine Nachkommastelle, gezeigt. Die Anzahl der Datenpunkte wird deutlich reduziert. Aus Tabelle 15 sind die konkreten Zahlen zu entnehmen. Es bleiben 109 Datenpunkte übrig, was einer Reduzierung um fast 2/3 entspricht. Außerdem ist zu erkennen, dass es auch vorkommen kann, dass zwei Punkte nach der Reduktion immer noch sehr nahe beieinander liegen können, siehe Abbildung 29 Mitte. Dies kann an der Grenze zwischen Auf- und Abrunden auftreten. In der dritten, unteren, Grafik in Abbildung 29 ist das Ergebnis der Reduktion mit dem Metrik-Verfahren und einem Schwellenwert von 0.1 dargestellt. (Die Gewichte wurden alle zu 1 gesetzt.) Es ist deutlich zu erkennen, dass mit diesem Verfahren noch stärker die Anzahl an Datenpunkten reduziert wurde. Es wurde um gut 4/5 reduziert auf 58 Datenpunkte. Es ist gut zu erkennen, dass es im Gegensatz zum Round-Verfahren deutliche Abstände zwischen den einzelnen Punkten gibt.

Rohdaten	300 Datenpunkte
Round (2 Dez.)	297 Datenpunkte
Round (1 Dez.)	109 Datenpunkte
Metrik (0.01)	282 Datenpunkte
Metrik (0.05)	135 Datenpunkte
Metrik (0.1)	58 Datenpunkte

Tabelle 15: Vergleich der Reduktionsverfahren und Schwellenwerte

In Abbildung 30 sind die Ergebnisse der Reduktion mit dem Metrik-Verfahren bei unterschiedlichen Schwellenwerten gezeigt. Die Rohdaten sind die gleichen 300 Zufallspunkte und die Gewichte sind ebenfalls alle auf 1 gesetzt.

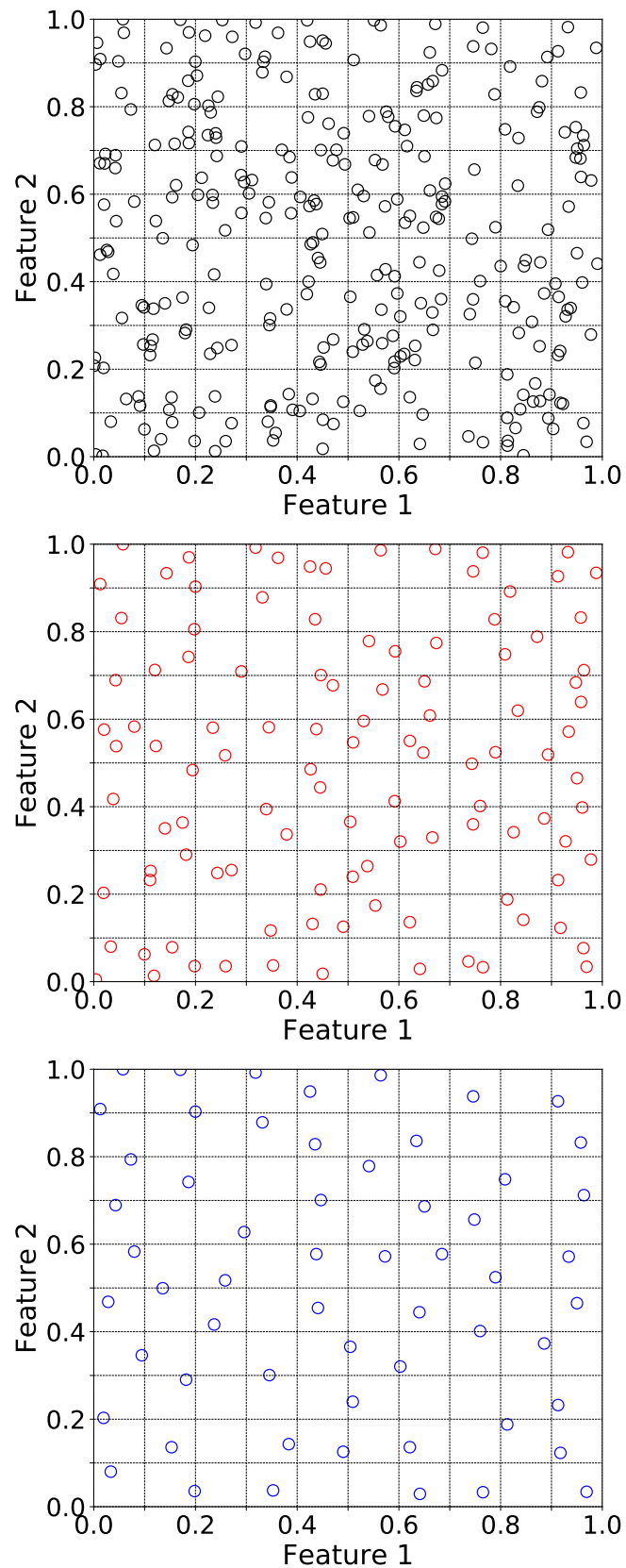


Abbildung 29: Vergleich der zufallsgenerierten Daten (schwarz, oben) mit den durch Verfahren Round, eine Nachkommastelle, (rot, Mitte) und Verfahren Metrik, Schwellenwert 0.01, (blau, unten) reduzierten Daten.

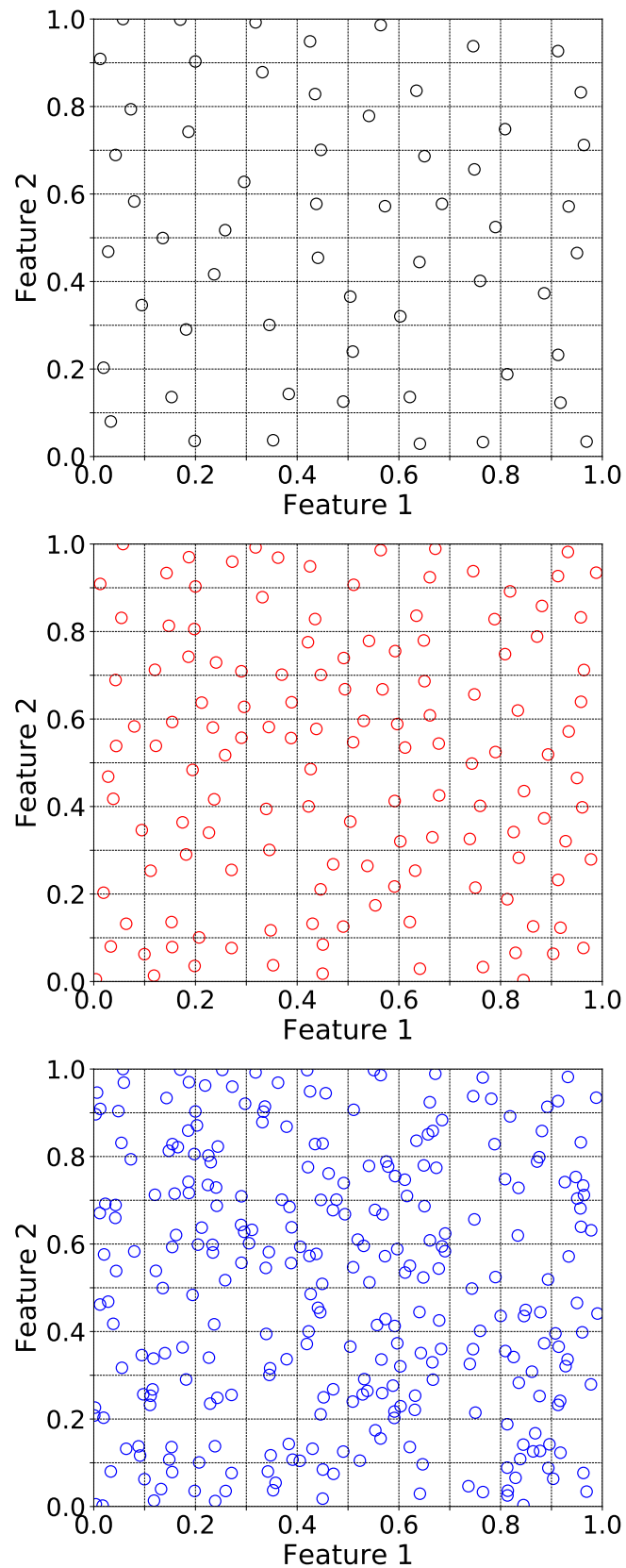


Abbildung 30: Vergleich der mit dem Verfahren Metrik reduzierten Daten mit unterschiedlichen Schwellenwerten: schwarz 0.1 (oben); rot 0.05 (Mitte); blau 0.01 (unten).

Die obere Grafik zeigt die, schon in Abbildung 29 zu sehende, Reduktion mit dem Schwellenwert 0.1. In der unteren Grafik ist die Reduzierung mit dem Schwellenwert 0.01 zu sehen. Es findet kaum eine Reduzierung statt. Wie aus Tabelle 15 zu entnehmen ist, werden gerade einmal 18 Datenpunkte entfernt. Die mittlere Grafik verdeutlicht den Vorteil des Metrik-Verfahrens variabel einstellbar zu sein, was die Stärke der Reduzierung angeht. Hier wird ein Schwellenwert von 0.05 verwendet, was zu einer Reduktion auf 135 Datenpunkte führt. So kann die Datenmenge durch die Wahl des Schwellenwertes angepasst werden, siehe hierzu auch die Werte für das Metrik-Verfahren in Tabelle 15. Ein Quervergleich des Round-Verfahrens (Abbildung 29 Mitte) mit dem Metrik-Verfahren mit Schwellenwert 0.05 (Abbildung 30 Mitte) zeigt bei vergleichbarer Datenpunktanzahl (109 bzw. 135) eine gleichmäßigere Verteilung beim Metrik-Verfahren.

In einem weiteren Schritt werden die beiden Verfahren an einem Datensatz, der durch das 1D-Modell erzeugt wird, getestet. Dafür wird die Parameterkombination $Pe = 50000$, $Da = 0.3$, $\tilde{c}_{B,0} = 5$ und $\partial_n \tilde{U}_n = 5$ gewählt. Außerdem werden $N_n = 331$ und $N_t = 10001$ gewählt.

Für das Metrik-Verfahren werden die Schwellenwerte 0.1, 0.07, 0.05 und 0.01 untersucht. Alle Gewichte werden auf 1 gesetzt. Bei dem Round-Verfahren wird auf eine, zwei und drei Nachkommastellen gerundet. Gegenüber gestellt wird der Grad der Reduktion in Form eines Vergleiches der Anzahl an Datenpunkten vor und nach der Reduktion. Außerdem wird die Dauer für den Reduktionsprozess verglichen. Hier ist zu beachten, dass die Reduktionszeit nur ein Anhaltspunkt sein kann. Es handelt sich nicht um eine gemittelte Zeit über mehrere Durchführungen, sondern um das Ergebnis eines Durchlaufs. Durch sonstige Einflüsse können Auswirkungen auf die Ausführungsdauer der jeweiligen Algorithmen aufgetreten sein, da auch parallel andere Prozesse auf dem Rechner durchgeführt wurden.

Verfahren	Datenpunkte	Zeit [sec]
Rohdaten	2740000	-
Round (3 Dez.)	973709	20
Round (2 Dez.)	162099	22
Round (1 Dez.)	19362	25
Metrik (0.01)	105808	5201
Metrik (0.05)	23892	1201
Metrik (0.07)	17403	876
Metrik (0.1)	12450	633

Tabelle 16: Vergleich der Reduktionsverfahren an realistischen Daten.

In Tabelle 16 sind die Ergebnisse zusammengefasst. Wie zu erkennen ist, ist das Metrik-Verfahren deutlich langsamer als das Round-Verfahren. Für vergleichbare Reduktionsschwellen reduziert das Metrik-Verfahren stärker, vergleiche Metrik (0.1) und Round (1 Dez.) oder Metrik (0.01) und Round (2 Dez.). Auch ist deutlich zu erkennen, dass es mit der variabel einstellbaren Schwelle des Metrik-Verfahrens möglich ist die Anzahl der Datenpunkte zu steuern.

Verfahren	Datenpunkte	Zeit [sec]
Rohdaten	2740000	-
Round (3 Dez.) + Metrik (0.01)	103710	1499
Round (3 Dez.) + Metrik (0.05)	23796	382
Round (3 Dez.) + Metrik (0.07)	17367	284
Round (3 Dez.) + Metrik (0.1)	12425	213

Tabelle 17: Kombination beider Reduktionsverfahren.

Eine Kombination aus beiden Verfahren ermöglicht es beide Vorteile miteinander zu verbinden. Die Schnelligkeit des Round-Verfahrens mit Einstellbarkeit an Datenpunkten des Metrik-Verfahrens. Dazu werden die Rohdaten zuerst mit dem Round-Verfahren reduziert (Runden auf drei Nachkommastellen) und anschließend der reduzierte Datensatz mit dem Metrik-Verfahren erneut reduziert. In Tabelle 17 sind die Ergebnisse dieser Kombination dargestellt. Der Vergleich mit Tabelle 16 zeigt, dass bei gleicher Schwelle die Anzahl an Datenpunkten nach der Reduktion ungefähr gleich groß ist, aber die Reduktionszeit um ungefähr den Faktor 3 reduziert wird.

5 Datenbasierte Ersatzmodelle

Es werden mehrere Datensätze betrachtet, mit denen verschiedene SGS-Modell-Varianten berechnet werden. Diese werden anschließend an einem einfachen Fall validiert. Der Validierungsfall ist gekennzeichnet durch $Pe = 50000$ und $Da = 0.3$. Die Validierung wird in Kapitel 6 ausführlich behandelt. Der erste Datensatz wird bezüglich der Péclet- und Damköhler-Zahl an den Validierungsfall angepasst. Bei Datensatz 2 wird der Pe -Zahlbereich so gewählt, dass der Validierungsfall enthalten ist, aber die Trainingsdaten nicht direkt mit diesen Parameterwerten generiert werden. Dies lässt einen Schluss auf die Verallgemeinerungsfähigkeiten des SGS-Modells zu. Bei beiden Datensätzen wird für die Reduktion mit dem Metrik-Verfahren ein Schwellwert von 0.05 angesetzt.

	Datensatz 1	Datensatz 2
Pe	50000	{45000, 60000}
Da	0.3	{0.1, 0.3}
$\tilde{c}_{B,0}$	{0.2, 1, 2, 3, 5}	{0.2, 1, 2, 3, 5}
$\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$	{-6, -4, -2, 0, 2, 5}	{-6, -4, -2, 0, 2, 5}

Tabelle 18: Parameterbereiche der unterschiedlichen Datensätze.

5.1 Ablauf

In den vorherigen Kapiteln wurden grundlegende Modellierungen und Funktionsweisen der Datengenerierung und Datenreduktion, die in dieser Arbeit verwendet werden, erörtert. Außerdem wurde die Arbeitsweise des SGS-Modells vorgestellt. Nun soll der Ablauf von der Datengenerierung bis zum fertigen SGS-Modell zusammengefasst werden.

- I) Zuerst müssen die Parameterbereiche unterteilt werden. Dies bedeutet, es wird festgelegt, mit wie vielen und welchen Péclet-Zahlen der Péclet-Zahlenbereich abgebildet wird. Das Gleiche wird für die Parameter Da , $\tilde{c}_{B,0}$ und $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ gemacht. Anschließend werden alle möglichen $(Pe, Da, \tilde{c}_{B,0}, \partial_{\tilde{n}}\tilde{U}_{\tilde{n}})$ -Kombinationen gebildet (es sei N_K die Anzahl an Kombinationen). Für jede der N_K Kombinationen werden mit dem 1D-Modell die Konzentrationsverläufe für die drei Spezies berechnet. Dabei ist zu beachten, dass der Berechnungsbereich bezüglich des Abstandes von der Grenzfläche mit $\tilde{L} = 0.03$ gewählt wird. Dies ist größer gewählt als der betrachtete Bereich von $\tilde{L}_{SGS} = 0.025$. Der Grund dafür ist, dass eventuelle Randeinflüsse vermieden werden sollen. Alle Daten für $\tilde{n} > 0.025$ werden nicht genutzt. Anschließend werden daraus alle Features und Label berechnet und die Datenpunkte generiert, siehe dazu Gleichung (32). Es ist zu beachten, dass die Datenpunkte in $(Pe, Da, \tilde{c}_{B,0}, \partial_{\tilde{n}}\tilde{U}_{\tilde{n}}, \tilde{l})$ -Einheiten für einen bestimmten Abstand \tilde{l} zur Grenzfläche gespeichert werden. Eine Einheit umfasst $N_t - 1$ Datenpunkte, da die Daten für $\tilde{l} = 0$ nicht benutzt werden (Anfangsbedingung). Alle Datenpunkte in den $N_K(N_t - 1)$ -Einheiten werden nun mit den globalen Minima und Maxima für jedes Feature und Label normiert, sodass jeder Eintrag eines Datenpunktes zwischen 0 und 1 liegt. Anschließend wird jede, der nun normierten, $N_K(N_t - 1)$ -Einheiten einer Reduktion unterzogen. Zuerst wird mit dem Round-Verfahren (Rundung auf 3. Nachkommastelle) vorreduziert, anschließend mit dem Metrik-Verfahren und der zu wählenden Schwelle nochmals reduziert. Danach werden alle reduzierten Einheiten zu einem Datensatz zusammengefasst.
- II) Als Nächstes wird dieser Datensatz analysiert und es werden die Features für das Netztraining ausgewählt. Die ausgewählten Features sollen die Daten möglichst gut beschreiben, aber die Anzahl an Features soll möglichst gering sein. Die Feature-Auswahl wird mit Hilfe von Korrelationsuntersuchungen und einem *Random Forest Regressor* getroffen.
- III) Dadurch dass Features vernachlässigt werden, wird der Datensatz auf einen Unterraum projiziert. Dies kann zu Häufung von Datenpunkten führen. Um diese herauszufiltern, wird wieder eine Reduktion mit Round- und Metrik-Verfahren durchgeführt.
- IV) Da für das Netztraining ein am Institut schon vorliegendes Template genutzt wird, muss der Datensatz nun noch auf die dafür notwendige Form konvertiert werden. Die Datenpunkte müssen wieder entnormalisiert werden und in dem richtigen Dateiformat gespeichert werden. Zu beachten ist, dass die Daten weiterhin dimensionslos sind, da die OpenFOAM-Simulationen daraufhin angepasst sind.
- V) Mit dem, an die Feature-Auswahl, angepassten Template wird nun für jedes Label ein eigenes neuronales Netz trainiert. Anschließend werden die Netze für Label einer Spezies zu je einem Modell zusammengefasst. Der Reaktionsterm wird isoliert behandelt. Dies bedeutet, dass vier Modelle erstellt werden, eines je Spezies A, B, P und eines für den Reaktionsterm. Diese bilden zusammen ein SGS-Modell.

5.2 Datenanalyse

Die vorherigen Kapitel haben ausführlich die Grundlagen für Schritt I) erläutert. Nun geht es um Schritt II), die Datenanalyse. Die reduzierten Trainingsdaten liegen vor und die Features sollen ausgewählt werden. Zuerst wird auf die Grundlagen der beiden Methoden (Korrelation und *Random Forest Regressor*) eingegangen, die hier als Entscheidungsgrundlage für die Feature-Auswahl dienen. Danach werden die Ergebnisse vorgestellt und die Feature-Auswahl getroffen. Darauf folgt das Netztraining.

5.2.1 Grundlagen

Korrelation

Eine einfache und schnelle Möglichkeit Zusammenhänge in einer Datenmenge zu erkennen ist die Korrelation zwischen zwei Merkmalen x und y zu berechnen. Hier wird der Korrelationskoeffizient $korr$ nach Bravais-Pearson betrachtet. Dieser ist ein maßstabsunabhängiges und dimensionsloses Zusammenhangsmaß. Er gibt den Grad eines linearen Zusammenhangs zwischen den Merkmalen wieder. Der Korrelationskoeffizient berechnet sich nach Gleichung (35). Für die Berechnung des Koeffizienten wird in dieser Arbeit die in *SciPy.stats* vorhandene Funktion benutzt. Der Korrelationskoeffizient kann Werte im Bereich $-1 \leq korr \leq 1$ annehmen. Bei $korr > 0$ liegt eine positive Korrelation vor. Mit steigendem x nimmt auch y zu. Analog dazu gibt es den Fall der negativen Korrelation ($korr < 0$), bei der y mit steigendem x abnimmt. Gilt $|korr| = 1$ spricht man von vollständiger Korrelation. Es liegt eine lineare Abhängigkeit zwischen den Merkmalen vor. Den Bereich $0 < |korr| < 0.5$ nennt man schwache, bzw. $0.5 \leq |korr| < 1$ starke Korrelation. Unkorreliertheit liegt bei $korr = 0$ vor. Dies bedeutet jedoch nicht, dass kein Zusammenhang zwischen den Merkmalen bestehen kann. Der Korrelationskoeffizient nach Bravais-Pearson lässt nur Aussagen über den linearen Zusammenhang zu [5].

$$korr = \frac{\overline{xy} - \bar{x} \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \sqrt{\overline{y^2} - \bar{y}^2}} \quad (35)$$

Random Forest Regressor

Als weiteres Hilfsmittel für die Feature-Auswahl wird ein *Random Forest Regressor* verwendet. Dieser zählt zu den Algorithmen des überwachten Lernens. Es handelt sich um eine Ensemble Methode. Dies bedeutet, dass das Ergebnis aus einer Gesamtbetrachtung vieler einzelner Algorithmen resultiert. Bei einem *Random Forest Regressor* sind die einzelnen Algorithmen *Decision Trees*, die immer individuell konstruiert werden und sich deshalb voneinander in ihrer Struktur unterscheiden. Diese individuellen Unterschiede in der Baum-Struktur können z.B. durch die Verwendung von *bootstrap*-Datensätzen beim Trainieren der Bäume erreicht werden. Ein *bootstrap*-Datensatz hat den gleichen Umfang wie der ursprüngliche Datensatz, wird jedoch zufällig aus den Datenpunkten zusammengestellt. Es ist demnach möglich, dass einzelne Datenpunkte mehrfach vorkommen. Für das Training jedes *Decision Trees* wird ein eigener *bootstrap*-Datensatz kreiert. So entstehen viele verschiedene Variationen an *Decision Trees*. Um den *Random Forest* für einen Datenpunkt auszuwerten, werden zunächst alle *Decision Trees* für diesen Punkt ausgewertet und die Ergebnisse werden gemittelt. Das Ergebnis eines *Random Forest* ist demnach die Mittelung über viele unterschiedliche *Decision Trees*. Dies ist von Vorteil, da ein einzelner *Decision Tree* zur Überanpassung an die Daten (*overfitting*) neigt. Dies bedeutet, dass die Trainingsdaten gut wiedergegeben werden, andere Daten jedoch nicht. Durch die Betrachtung vieler unterschiedlicher Bäume wird die Tendenz zum *overfitting* ausgeglichen [15].

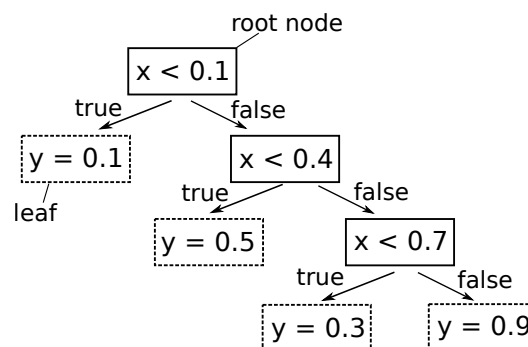


Abbildung 31: Skizze eines beispielhaften Decision Tree.

Decision Trees sind geordnete, gerichtete Bäume, welche sowohl für Klassifikations- als auch Regressionsaufgaben benutzt werden können. Für Regressionsaufgaben haben die Endknoten, auch *leaf* genannt, numerische Werte. Um den Wert eines Punktes mit Hilfe eines Baumes vorherzusagen, folgt man, ausgehend vom Wurzel-Knoten, einem Pfad zu einem Endknoten. An jedem Knoten wird eine Entscheidung über den weiteren Weg anhand eines Kriteriums (z.B. $>$ oder $<$ als ein bestimmter Wert) getroffen. Dies ist vergleichbar mit if-Bedingungen in Programmiersprachen. In Abbildung 31 ist ein beispielhafter *Decision Tree* skizziert. Die Kriterien eines Knoten werden durch das Training des *Decision Trees* festgelegt. Eine anschauliche Erklärung des Trainingvorganges ist als Video in [16] gegeben. Um zu verhindern, dass für jeden Trainingsdatenpunkt ein eigener Endknoten gebildet wird (*overfitting*) und somit keine Generalisierung mehr stattfindet, gibt es einige Möglichkeiten, z.B. kann die Tiefe des Baums ("max depth") beschränkt werden oder die Mindestanzahl an Datenpunkten ("min split"), um einen Verzweigungsknoten zu erzeugen (siehe [15]).

Für die Bestimmung solcher Parameter wird eine Hyperparameter-Optimierung für den Random Forest Regressor durchgeführt. Dafür wird Datensatz 2 benutzt. Variiert werden dabei die Parameter "max depth", "min split" und "Anzahl Features". Die Anzahl der Bäume wird auf 400 gesetzt. Die Hyperparameter-Optimierung wird mit Hilfe des *sklearn model selection*-Tool *GridSearchCV* durchgeführt, siehe [17]. Als Auswahlmetrik dient die *score*-Funktion des *Random Forest Regressor*. Die Ergebnisse der Optimierung sind in Tabelle 19 aufgelistet. Für die Anzahl an Features ("n feature") werden die Optionen 'auto' und 'sqrt' getestet.

Label	"max depth"	"min split"	"n feature"
$\partial_{\tilde{n}} \tilde{c}_{A,\Sigma}$	30	2	'auto'
$\partial_{\tilde{n}} \tilde{c}_A$	30	2	'auto'
\tilde{c}_A	30	2	'auto'
$\partial_{\tilde{n}} \tilde{c}_B$	20	15	'auto'
\tilde{c}_B	30	2	'auto'
$\partial_{\tilde{n}} \tilde{c}_P$	30	10	'auto'
\tilde{c}_P	30	2	'auto'
R	30	2	'auto'

Tabelle 19: Verwendete Hyperparameter für den *Random Forest Regressor*.

Diese Parameterwerte werden für alle im Weiteren angewandeten *Random Forest Regressoren* verwendet.

5.2.2 Feature-Auswahl

Für das Netztraining und die Implementierung in OpenFOAM werden vorgefertigte Vorlagen verwendet [18]. Diese sehen vor, dass zwar für jedes Label ein eigenes Netz trainiert wird (Netze mit je einer Ausgabe), aber die Netze der Label für jede Spezies zusammengefasst werden. Dies bedeutet, dass die Netze für die Label der Spezies A $\partial_{\tilde{n}} \tilde{c}_A|_{\Sigma}$, $\partial_{\tilde{n}} \tilde{c}_A|_{\tilde{l}}$ und $\tilde{c}_A|_{\tilde{l}}$ zu einem Modell zusammengefasst werden. Das Gleiche gilt für Spezies B mit den Label $\partial_{\tilde{n}} \tilde{c}_B|_{\tilde{l}}$, $\tilde{c}_B|_{\tilde{l}}$ und Spezies P mit $\partial_{\tilde{n}} \tilde{c}_P|_{\tilde{l}}$, $\tilde{c}_P|_{\tilde{l}}$. Der Reaktionsterm ist ein eigenes Modell für sich. Daraus folgt, dass es für jede Simulation vier Modelle gibt, die zusammen ein SGS-Modell bilden. Die neuronalen Netze für jedes Label einer Spezies müssen die gleichen Eingabegrößen haben. Die Modelle der verschiedenen Spezies können jedoch verschiedene Eingabegrößen haben. Dies bedeutet, dass z.B. die *Feature Importance* für die Label der Spezies A gemeinsam betrachtet werden müssen. Im Folgenden werden nun für jeden Datensatz die Korrelationen der Feature/Label berechnet und die *Feature Importance* per *Random Forest Regressor* bestimmt. Letztere ist eine Methode des Random Forest. Wichtigere Features haben einen höheren Wert als Unbedeutendere.

Datensatz 1

Zuerst wird die Korrelation zwischen den einzelnen Features/Label betrachtet. Dies kann schon erste Zusammenhänge aufzeigen. Der Übersichtlichkeit halber wird an dieser Stelle nicht die komplette Korrelationsmatrix zwischen allen Features und Label gezeigt, sondern nur die Korrelationen der Features untereinander (Abbildung 32 links) und der Korrelationen der Features mit den Label (Abbildung 32 rechts). Da dieser Datensatz nur auf einer Péclet-Zahl und einer Damköhler-Zahl beruht, werden diese nicht mit untersucht. Hierbei ist ebenso zu beachten, dass die Korrelationskoeffizienten auf zwei Nachkommastellen gerundet sind.

Wie in der linken Grafik in Abbildung 32 zu erkennen ist, liegt eine starke Korrelation zwischen der vorgegebenen Anfangs-/Randwertbedingung $\tilde{c}_{B,0}$ und der mittleren Konzentration \tilde{c}_B bis zu einem Abstand \tilde{l} von der Grenzfläche vor. Ein Zusammenhang zwischen diesen Größen ist zu erwarten, da $\tilde{c}_{B,0}$ eine obere Schranke für \tilde{c}_B darstellt. Spezies B kann nicht zusätzlich produziert werden und je höher die Anfangskonzentration desto höher sind auch die mittleren

Konzentrationen. Auch ist zu erkennen, dass eine Korrelation zwischen der mittleren Konzentration von Spezies A und Spezies P vorliegt. Bei den Korrelationen der Features mit den Label (rechts) fällt auf, dass es keine starke Korrelationen der Features $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ und \tilde{n} mit allen Label gibt. \tilde{n} gibt den Abstand zur Grenzfläche an und $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ lässt sich als der indirekte Polarwinkel Θ interpretieren, siehe Abbildung 23. Es liegt demnach zumindest kein linearer Zusammenhang zwischen dem Ort (um eine Blase) und den Label vor. Es ist auch zu erkennen, dass für jede Spezies eine starke Korrelation zwischen den mittleren Konzentrationen und der Konzentration bei Abstand \tilde{l} vorliegt.

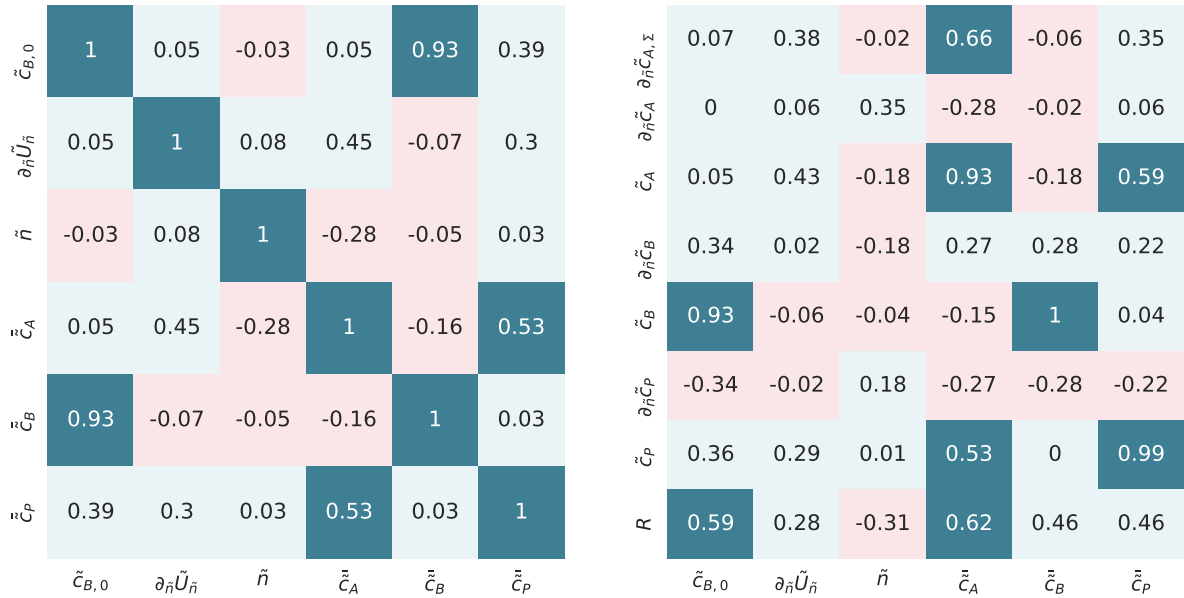


Abbildung 32: Bravais-Pearson Korrelationskoeffizienten zwischen Features (links), Featuren und Label (rechts) des Datensatz 1.

In Abbildung 33 sind die Ergebnisse der *Feature Importance* des *Random Forest Regressors* für jedes Label abgebildet. Wie zu erkennen ist, haben die Péclet-Zahl und die Damköhler-Zahl keinen Einfluss. Dies ist logisch, da diese nicht variiert werden.

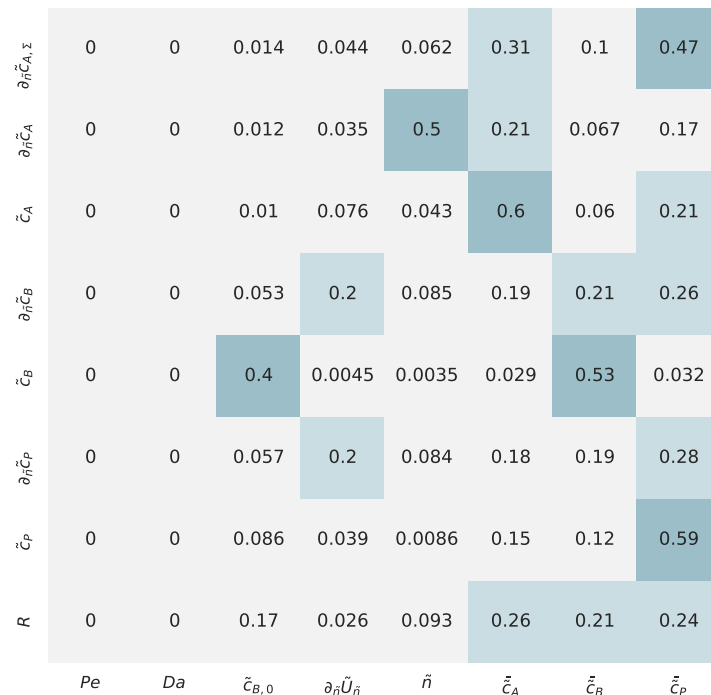


Abbildung 33: "Feature Importance" des Random Forest Regressors für die Label des Datensatz 1.

Betrachtet man die in Abbildung 33 dargestellten Ergebnisse sind viele Gemeinsamkeiten mit den Korrelationen aus Abbildung 32 (rechts) zu erkennen. Für die Label der Spezies A ist zu erkennen, dass für diese die Features \tilde{n} , \tilde{c}_A und \tilde{c}_P von Bedeutung sind. Die Bedeutung der mittleren Konzentration von Spezies P als Feature lässt sich mit der Korrelation von \tilde{c}_A und \tilde{c}_P erklären. Diese drei Features werden als Features für die Spezies A behalten, die anderen Features werden vernachlässigt. Für Spezies B sind nach Abbildung 33 die Features $\tilde{c}_{B,0}$, $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$, \tilde{c}_B und \tilde{c}_P von Bedeutung. $\tilde{c}_{B,0}$ ist ein globaler Parameter. Dieser soll wenn möglich als Feature vermieden werden, da so eine breitere Anwendung des Modells möglich ist. Aufgrund der sehr starken Korrelation mit der mittleren Konzentration wird dessen Einfluss mit \tilde{c}_B mit abgebildet. Dementsprechend wird $\tilde{c}_{B,0}$ als Feature verworfen. Außerdem gibt der *Random Forest* $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ als wichtiges Feature an. In Arbeiten mit vergleichbaren Fällen, [10] und [12], wurde der Polarwinkel Θ stets vernachlässigt. $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ ist quasi eine indirekte Beschreibung des Polarwinkels, siehe Abbildung 23. Es werden zwei verschiedene Kombinationen an Features betrachtet. In Version α wird $\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$ als Feature vernachlässigt, in Version β als Feature benutzt, dies gilt ebenso für Spezies P. Außerdem wird in beiden Versionen \tilde{n} als Feature hinzugefügt, obwohl es nach Abbildung 33 für Spezies B vernachlässigbar ist. Der Grund dafür ist, dass nur mit der mittleren Konzentration das Konzentrationsprofil nicht rekonstruiert werden kann. Dies lässt sich an einem Beispiel verdeutlichen. So kann bei einem großen Abstand mit einem flachen Konzentrationsprofil die gleiche gemittelte Konzentration erreicht werden, wie bei einem kleinen Abstand und einem fülligeren Profil. Für Spezies P und das Reaktionsterm-Label R wird ebenso \tilde{n} als Feature verwendet. Für Spezies P ist ansonsten nur das Feature \tilde{c}_P von Bedeutung. Der Reaktionsterm wird neben \tilde{n} die Features \tilde{c}_A , \tilde{c}_B und \tilde{c}_P haben.

In Tabelle 20 sind die unterschiedlichen Feature-Kombinationen für Datensatz 1 aufgelistet.

Spezies	Version α	Version β
A	$\tilde{n}, \tilde{c}_A, \tilde{c}_P$	$\tilde{n}, \tilde{c}_A, \tilde{c}_P$
B	$\tilde{n}, \tilde{c}_B, \tilde{c}_P$	$\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}, \tilde{n}, \tilde{c}_B, \tilde{c}_P$
P	\tilde{n}, \tilde{c}_P	$\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}, \tilde{n}, \tilde{c}_P$
R	$\tilde{n}, \tilde{c}_A, \tilde{c}_B, \tilde{c}_P$	$\tilde{n}, \tilde{c}_A, \tilde{c}_B, \tilde{c}_P$

Tabelle 20: Verwendete Feature Kombinationen für Datensatz 1 und 2.

Datensatz 2

Für Datensatz 2 werden ebenfalls zuerst die Korrelationen zwischen den einzelnen Features (Abbildung 34 links) und die Korrelationen zwischen den Features und Label (Abbildung 34 rechts) betrachtet. Da in diesem Datensatz mehr als nur eine *Pe*- und *Da*-Zahl eingehen, werden auch für diese Größen die Korrelationen betrachtet.

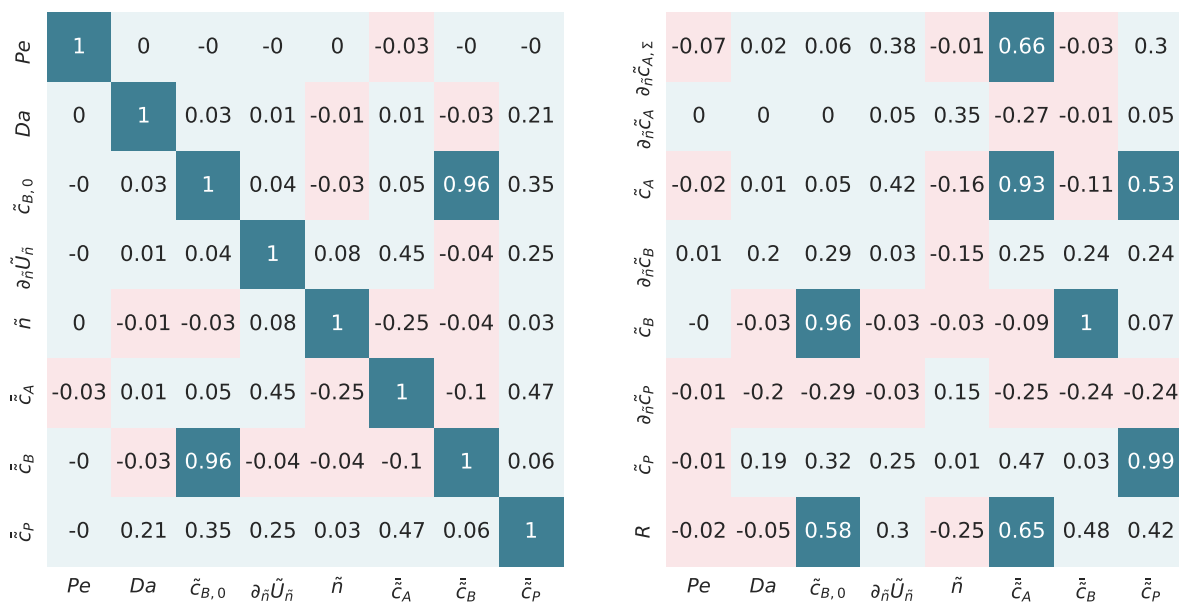


Abbildung 34: Bravais-Pearson Korrelationskoeffizienten zwischen Features (links), Features und Label (rechts) des Datensatz 2.

Analog zum Datensatz 1 sind in Abbildung 34 die Korrelationskoeffizienten für Datensatz 2 abgebildet. Es ergeben sich im Großen und Ganzen die gleichen Zusammenhänge wie schon bei Datensatz 1, weswegen auf diese Aspekte an dieser Stelle nicht noch einmal eingegangen wird. An dieser Stelle soll nur noch kurz auf die Péclet- und Damköhler-Zahl eingegangen werden, da diese im vorherigen Datensatz nicht behandelt wurden. Es ist deutlich zu erkennen, dass diese beiden Größen keine nennenswerten Korrelationen mit anderen Features oder Label aufweisen. Dies kann jedoch bei einem größeren Variationsbereich und einer größeren Anzahl von Werten für die Pe - und Da -Zahl anders sein. Zur Erinnerung, im Datensatz 2 werden je nur zwei verschiedene Pe - bzw. Da -Zahlen verwendet.

Für Datensatz 2 werden ebenfalls die Ergebnisse der *Feature Importance* des *Random Forest Regressors* betrachtet. Diese sind in Abbildung 35 dargestellt.

$\partial_{\tilde{n}}\tilde{c}_{A,\Sigma}$	0.0038	0.0035	0	0.0028	0.12	0.19	0.024	0.65
$\partial_{\tilde{n}}\tilde{c}_A$	0.0003	0.0007	0	0.0002	0.68	0.2	0.0015	0.11
\tilde{c}_A	0	0	0	0.0008	0.0007	0.99	0.0001	0.0068
$\partial_{\tilde{n}}\tilde{c}_B$	0.0024	0.013	0.0074	0.27	0.099	0.13	0.025	0.46
\tilde{c}_B	0	0	0	0	0	0.0008	1	0.0018
$\partial_{\tilde{n}}\tilde{c}_P$	0.0026	0.015	0.0063	0.27	0.097	0.13	0.028	0.46
\tilde{c}_P	0	0	0	0.0001	0	0.0091	0.0013	0.99
R	0	0	0.036	0	0.057	0.16	0.29	0.45
	Pe	Da	$\tilde{c}_{B,0}$	$\partial_{\tilde{n}}\tilde{U}_{\tilde{n}}$	\tilde{n}	\tilde{c}_A	\tilde{c}_B	\tilde{c}_P

Abbildung 35: "Feature Importance" des Random Forest Regressors für die Label des Datensatz 2.

Es ergibt sich ein ähnliches Bild wie bei Datensatz 1. Die gleichen Feature-Kombinationen wie für Datensatz 1 werden gewählt. Dies geschieht mit den gleichen Begründungen für Wegfall oder Hinzunahme eines Features, welches nach Abbildung 35 eine größere bzw. kleinere Bedeutung hat. Es soll noch kurz auf \tilde{c}_A als Feature für das Reaktionsterm-Label eingegangen werden. Nach der *Feature Importance* spielt es eine untergeordnete Bedeutung, jedoch liegt eine starke Korrelation mit R vor, siehe Abbildung 34 rechts. Dies dient als Begründung es als Feature aufzunehmen und somit die gleichen Kombinationen wie bei Datensatz 1 zu haben.

5.3 Netztraining

Nach der Wahl der Features werden die Netze für die einzelnen Datensätze trainiert. Wie schon in Kapitel 2.5 beschrieben, können neuronale Netze variabel gestaltet werden. Es muss also eine Wahl getroffen werden, aus wie vielen versteckten Schichten das Netz bestehen und wie viele Neuronen es pro Schicht haben soll. Diese Parameter haben Einfluss auf die Regression durch das Netz. Um die optimalen Einstellungen für ein solches Netz zu finden, gilt es eine Hyperparameteroptimierung durchzuführen. Aus Zeitgründen wird in dieser Arbeit auf eine solche Optimierung verzichtet. Es werden Hyperparameterwerte aus der Literatur entnommen. In [12] werden ähnliche Fälle betrachtet, also bietet es sich an die Hyperparameter von dort zu übernehmen. Die neuronalen Netze haben dementsprechend 4 versteckte Schichten und 50 Neuronen pro versteckter Schicht. Als Aktivierungsfunktion wird eine *Rectified Linear Unit* (ReLU) verwendet, welche eine Rampenfunktion darstellt. In [12] wird auch auf unterschiedliche Aktivierungsfunktionen eingegangen. Die Modelle werden für 5000 Epochen trainiert. Für das Modelltraining gibt es eine Jupyter Notebook Vorlage unter [18]. Diese wird verwendet und für den jeweiligen Fall angepasst.

Datensatz	Anzahl Datenpunkte	Ø Trainingszeit pro Label
Set 1- α	10383	2 min 40 sec
Set 1- β	22932	4 min 15 sec
Set 2- α	14990	3 min 03 sec
Set 2- β	31527	5 min 39 sec

Tabelle 21: Datensatzgröße und Trainingszeiten für die unterschiedlichen Datensatz-Versionen.

Wie im vorherigen Abschnitt beschrieben werden für die einzelnen Datensätze mehrere Feature-Kombinationen betrachtet. Es werden einige mögliche Features vernachlässigt, dementsprechend wird für die unterschiedlichen Kombinationen jeweils eine weitere Reduktion der Daten durchgeführt, vergleiche Schritt III) in Kapitel 5.1. In der zweiten Spalte von Tabelle 21 ist die Anzahl an Datenpunkten für jede Datensatz-Version aufgelistet. Mit dieser Anzahl an Datenpunkten wird jeweils das Netztraining durchgeführt.

An dieser Stelle erfolgt nun ein kleiner Einschub. Es wird kurz angesprochen wie sich die Anzahl der Datenpunkte über Schritt I) bis III) entwickelt hat. Datensatz 1 umfasst nach der Datengenerierung 41,100,000 Datenpunkte. Durch die erste Reduktion, in Schritt I) enthalten, verringert sich die Anzahl auf 119,990 Datenpunkte. Mit dieser Anzahl wurde die Datensatzanalyse in Schritt II) durchgeführt. Für Datensatz 2 lagen ursprünglich 164,400,000 Datenpunkte vor. Die erste Reduktion verringerte diese auf 451,340 Datenpunkte. Durch die zweite Reduktion, Schritt III), verringert sich die Datenmenge der jeweiligen Datensatz-Versionen auf die in Tabelle 21 genannten Werte. Für Datensatz 1 entspricht dies einer Reduktion der Datenmenge um einen Faktor > 1700 und für Datensatz 2 um einen Faktor > 5200 .

Die dritte Spalte in Tabelle 21 zeigt die durchschnittliche Trainingszeit pro Label. Es ist deutlich zu erkennen, dass mit steigender Anzahl an Datenpunkten bei gleicher Feature-Kombination die Trainingszeit steigt. So zeigt ein Vergleich von "Set 1- α " mit "Set 2- α " eine Vergrößerung der Datenmenge um den Faktor 1.44, woraus ein Anstieg der Trainingszeit um den Faktor 1.14 erfolgt. Ein Vergleich zwischen "Set 1- β " und "Set 2- β " zeigt einen noch größeren Anstieg der Trainingszeit. Um den Anstieg der Trainingszeit zu verdeutlichen wird ein Netztraining mit Datensatz 1 mit Feature-Kombination α durchgeführt. Bei dieser Variante entfällt jedoch die zweite Reduktion, sodass mit 119,990 Datenpunkten trainiert wird. Es ergibt sich eine durchschnittliche Trainingsdauer pro Label von 27 Minuten und 12 Sekunden. Das Netztraining dauert demnach mehr als zehnmal so lange wie für "Set 1- α ". Die hier gezeigten Werte sind nur Stichproben und parallel laufende Prozesse können die Trainingszeiten beeinflusst haben, jedoch lässt sich eine Tendenz herauslesen. Mit steigender Datenmenge nimmt die Trainingszeit zu. Dies verdeutlicht die Notwendigkeit der Datenreduktion. Mit den unreduzierten Datensätzen sind die Trainingszeiten deutlich länger. Außerdem muss mehr Speicherkapazität aufgewendet werden.

Um Verwirrungen zu vermeiden wird darauf hingewiesen, dass im Folgenden unterschieden wird zwischen "Set 1- α " und "Modell 1- α ". Mit "Modell 1- α " wird die SGS-Modell-Variante bezeichnet, welche mit Datensatz "Set 1- α " trainiert wurde. Analoge Bezeichnungen gelten für die anderen Datensätze und deren SGS-Modelle.

Als Nächstes folgt eine Betrachtung der Fehler der berechneten SGS-Modell-Varianten. Hierfür wird eine Notebook-Vorlage aus [19] benutzt. Die Fehlerbetrachtung geschieht anhand der Abweichungen zwischen den Label-Werten der Trainingsdaten und den dazugehörigen vom Modell berechneten Label-Werten (mit $^+$ gekennzeichnet). Außerdem sind alle Label mit einer Minimum-Maximum-Skalierung auf den Wertebereich $[0, 1]$ skaliert. Dies ist durch das Subskript 01 kenntlich gemacht.

An dieser Stelle wird nur auf eines der vier Modelle eingegangen. Die Diagramme für die restlichen Modelle werden im Anhang F aufgeführt. Gewählt wird "Modell 2- α ". Dies hat den Vorteil, dass das Modell nicht nur an den eigenen

Trainingsdaten, “Set 2- α ”, getestet werden kann, sondern auch an Daten, die nicht dafür verwendet wurden. Dies sind die Daten des Datensatzes “Set 1- α ”. Die Parameter-Kombination, mit der “Set 1- α ” generiert wurde, liegt innerhalb des Parameterbereichs für den “Modell 2- α ” trainiert wurde.

In der Abbildung 36 sind die Histogramme für die Label der Spezies A dargestellt. Diese zeigen die Abweichungen zwischen den Trainingsdaten “Set 2- α ” und den von “Modell 2- α ” berechneten Werten. Wie zu erkennen ist, haben die Abweichungen der drei Label die gleiche Größenordnung. Ein Vergleich mit den Abweichungen bezüglich des Datensatzes “Set 1- α ” zeigt, dass auch in diesem Fall ähnliche Größenordnungen gelten, mit Ausnahme der Normalableitung an der Grenzfläche, siehe Abbildung 37. Für “Set 1- α ” tendiert das Modell zu einer Überschätzung der Label-Werte.

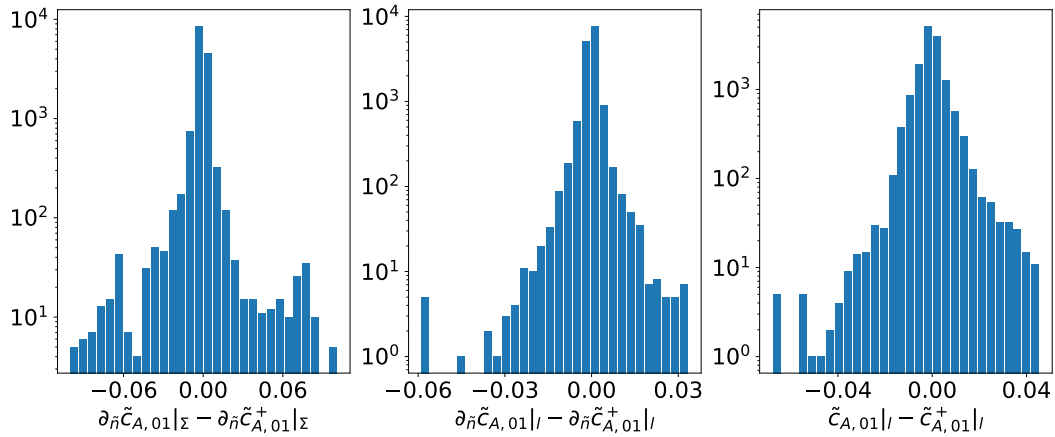


Abbildung 36: Histogramm für Abweichungen der Label-Werte des Trainingsdatensatzes gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies A.

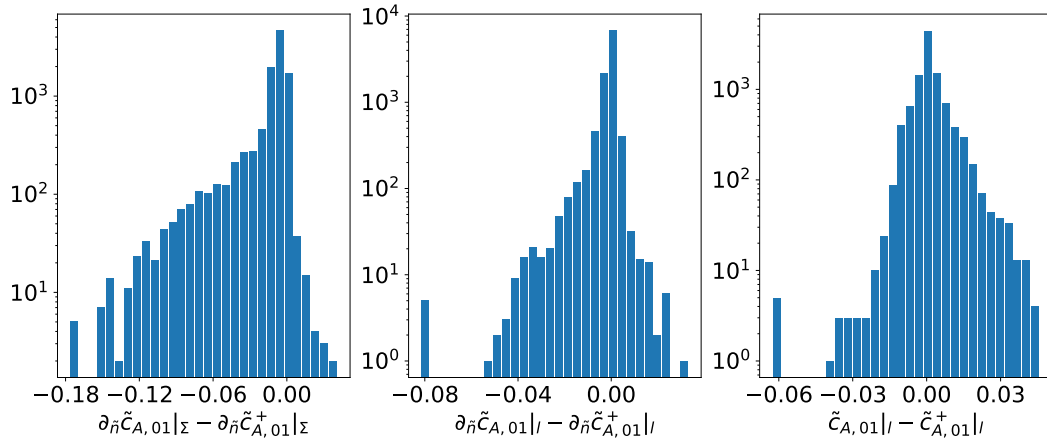


Abbildung 37: Histogramm für Abweichungen der Label-Werte des Testdatensatzes, “Set 1- α ”, gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies A.

Analog zu Spezies A sind in den Abbildungen 38 und 39 die Histogramme für die Label der Spezies B und P abgebildet. Wie zu erkennen ist, sind in beiden Abbildungen die Abweichungen der Normalableitungen der Konzentrationen deutlich größer als bei den anderen Label, auch im Vergleich zur Spezies A. Es ist keine eindeutige Tendenz herauszulesen, ob der Wert eines Label von dem Modell über- oder unterschätzt wird. Die Label der Spezies B werden in beiden Datensätzen eher unterschätzt. Die Label der Spezies P hingegen eher überschätzt.

In Abbildung 40 sind die Histogramme für den Reaktionsterm abgebildet. Die linke Grafik beschreibt die Abweichungen für den Trainingsdatensatz “Set 2- α ”. Die rechte Grafik stellt die Abweichungen für “Set 1- α ” dar. Wie zu erkennen ist, schätzt das SGS-Modell die Trainingsdaten (links) besser ab. Die Abweichungen sind etwa halb so groß wie für “Set 1- α ”, aber in beiden Fällen sind die Abweichungen klein. Die Label-Werte für “Set 1- α ” werden von dem Modell eher unterschätzt.

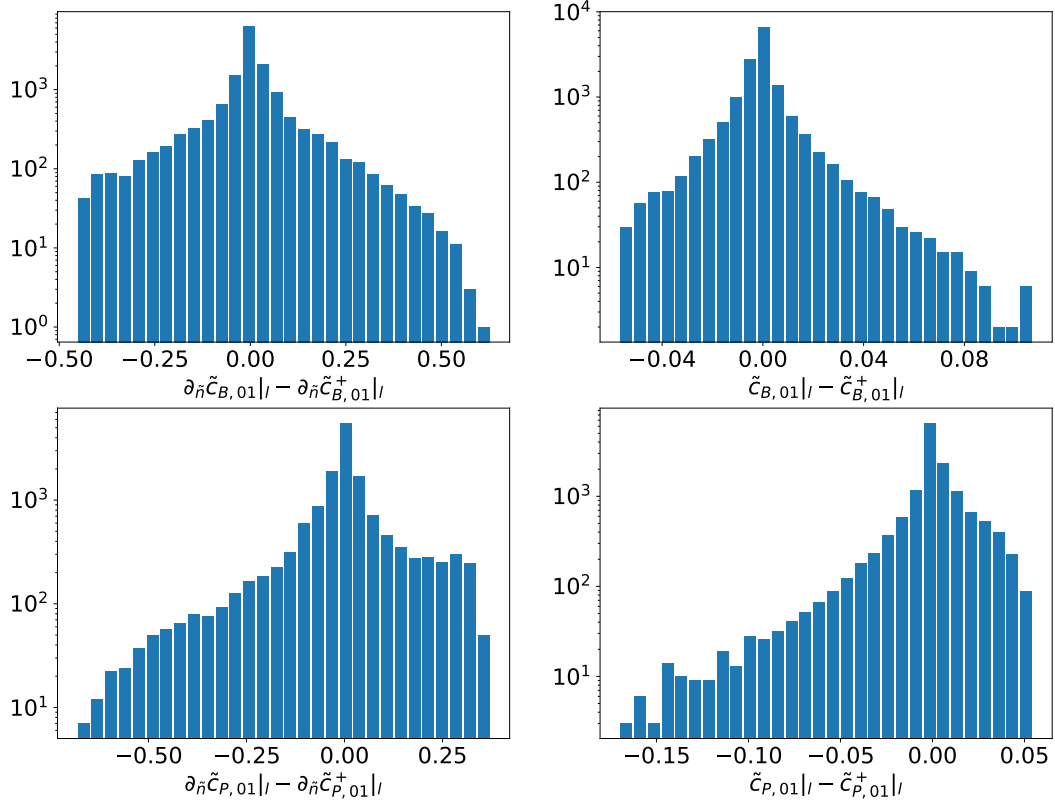


Abbildung 38: Histogramm für Abweichungen der Label-Werte des Trainingsdatensatzes gegenüber den von "Modell 2- α " berechneten Label-Werten $^+$ für die Label der Spezies B und P.

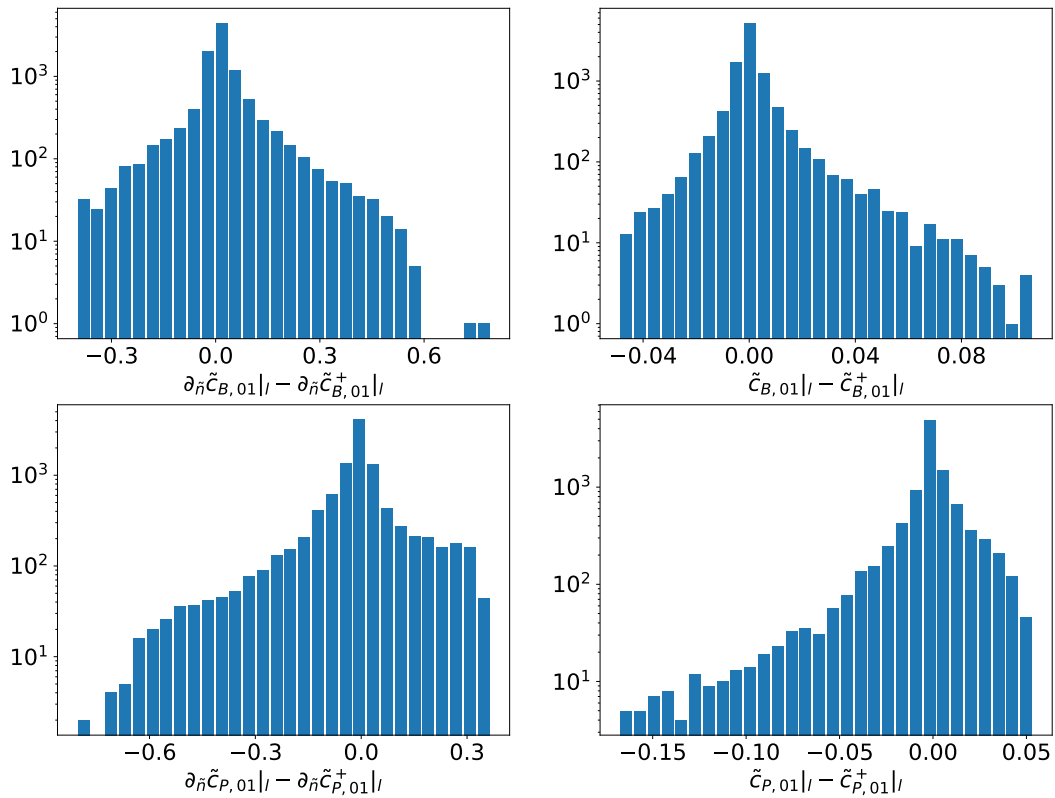


Abbildung 39: Histogramm für Abweichungen der Label-Werte des Testdatensatzes, "Set 1- α ", gegenüber den von "Modell 2- α " berechneten Label-Werten $^+$ für die Label der Spezies B und P.

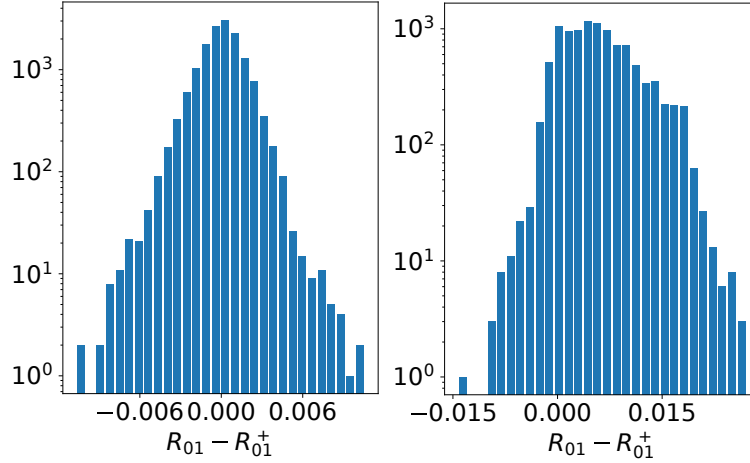


Abbildung 40: Histogramm für Abweichungen der Label-Werte der Datensätze gegenüber den von “Modell 2- α ” berechneten Label-Werten R^+ für das Label Reaktionsterm. Links für “Set 2- α ”, rechts “Set 1- α ”.

Die Label $\partial_{\tilde{n}} \tilde{c}_A|_{\Sigma}$, $\partial_{\tilde{n}} \tilde{c}_A|_{\tilde{l}}$, $\tilde{c}_A|_{\tilde{l}}$, $\tilde{c}_B|_{\tilde{l}}$, $\tilde{c}_P|_{\tilde{l}}$ und $R|_{\tilde{l}}$ schätzt das SGS-Modell für beide Testdatensätze mit kleinen bis moderaten Abweichungen ab. Aber die Werte der Normalableitungen der Konzentrationen für Spezies B und P werden in beiden Fällen sehr schlecht abgeschätzt. Abweichungen von bis zu ± 0.6 bei einem Wertebereich von $[0, 1]$ sind sehr groß.

Diese großen Abweichungen können daran liegen, dass das Feature $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ vernachlässigt wurde. Der *Feature Importance* nach ist dieses Feature jedoch für die Label $\partial_{\tilde{n}} \tilde{c}_B|_{\tilde{l}}$ und $\partial_{\tilde{n}} \tilde{c}_P|_{\tilde{l}}$ von Bedeutung, vergleiche Abbildung 35. Ein Blick auf die Histogramme, siehe Anhang F, für “Modell 1- β ” und “Modell 2- β ” verstärken diese Vermutung. Beide benutzen $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ als Feature. Für “Modell 2- β ” sind die Abweichungen geringer, jedoch mit 30% immer noch recht groß. Bei “Modell 1- β ” liegen zumindest für $\partial_{\tilde{n}} \tilde{c}_B|_{\tilde{l}}$ erhebliche geringere Abweichungen vor.

6 Validierung

Validierungsfall

Die Validierung des SGS-Modells wird an einem Testfall vorgenommen, der in [20] als Vorlage vorbereitet ist. Es handelt sich um den Fall “water-air mit $d_b = 1\text{ mm}$ ”. Die Form der Blase ist einer zweiphasigen Simulation entnommen und kommt einer sphärischen Blase sehr nahe. Die Blase hat eine feste äußere Kontur und der Durchmesser entspricht $d_{eq} = 1\text{ mm}$. Die Form der Blase ist der des in dieser Arbeit verwendeten Ersatzproblems (von dem das 1D-Modell abgeleitet wurde) sehr ähnlich. Die Simulationsparameter sind $Pe = 50000$, $Da = 0.3$ und $c_{B,0} = 1$, dabei ergibt sich die Pe -Zahl aus $Re = 250$ und $Sc = 200$.

Vorgehen

Es werden vier unterschiedlich fein aufgelöste Gitter verwendet, wobei “Mesh 0” das gröbste und “Mesh 3” das feinste Gitter ist. “Mesh 3” hat dabei eine Auflösung, die fein genug ist, um den Stofftransport mit der standardmäßigen linearen Interpolation genau zu berechnen. Dementsprechend wird die Simulation auf diesem Gitter ohne SGS-Modell aber mit linearer Interpolation als Referenz zur Bestimmung der Güte der Simulationen benutzt. Die Simulationen gestalten sich wie folgt. Die Berechnung des Stofftransports und der Strömungsdynamik werden entkoppelt. Zuerst wird auf jedem Gitter das Geschwindigkeits- und Druckfeld bis zum stationären Zustand berechnet. Die stationären Geschwindigkeitsfelder werden anschließend für die Berechnung des Stofftransportes benutzt. Dies geschieht auf zwei Arten. Einmal wird auf jedem Gitter der Stofftransport mit Korrektur durch das SGS-Modell berechnet. Zusätzlich werden auf jedem Gitter Simulationen nur mit der linearen Interpolation ohne Korrektur durch das SGS-Modell durchgeführt. Dies dient als Vergleich für die Ergebnisse mit SGS-Modell.

Die Leistung des Subgridskalen-Modells wird anhand der globalen Sherwood-Zahl Sh bezüglich der Transferspezies A bewertet. Diese ist ein globaler Kennwert und ermöglicht somit einen einfachen und schnellen Vergleich auf den unterschiedlich feinen Gittern zwischen den Simulationen mit und ohne SGS-Modell. Die Verwendung eines globalen Kennwerts schließt jedoch nicht aus, dass selbst bei guter Übereinstimmung der globalen Sherwood-Zahl, es zu Abweichungen bezüglich lokaler Größen kommen kann. So kann es Abweichungen bezüglich der lokalen Sherwood-Zahl geben, die sich im Mittel aufheben.

Ergebnisse

In Tabelle 22 sind die Ergebnisse für diesen Testfall aufgelistet. Es werden sowohl die globalen Sherwood-Zahlen genannt, als auch die prozentuale Abweichung dieser zur Sh -Zahl der Referenzsimulation (“linear, Mesh 3”).

Simulation	Mesh 0	Mesh 1	Mesh 2	Mesh 3	Ø Abweichung
linear	175.72 (−32.52%)	241.11 (−7.41%)	259.36 (−0.40%)	260.41	
Modell 1- α	264.84 (+1.70%)	268.22 (+3.00%)	269.05 (+3.32%)	260.66 (+0.10%)	2.03 %
Modell 1- β	261.08 (+0.26%)	266.62 (+2.39%)	256.84 (−1.37%)	261.39 (+0.38%)	1.10 %
Modell 2- α	264.37 (+1.52%)	261.58 (+0.45%)	260.86 (+0.17%)	261.94 (+0.59%)	0.68 %
Modell 2- β	266.66 (+2.40%)	262.23 (+0.70%)	263.26 (+1.09%)	260.43 (+0.01%)	1.05 %

Tabelle 22: Globale Sherwood-Zahlen für die unterschiedlichen Simulationen und unterschiedliche Gitter.

An den globalen Sh -Zahlen für die lineare Interpolation ohne SGS-Modell kann man sehr gut erkennen, welche Auswirkungen eine unzureichende Auflösung hat. So wird bei den Gittern “Mesh 0” bis “Mesh 2” der Stoffübergang von Spezies A unterschätzt. Für das sehr grobe Gitter sogar mit einer Ungenauigkeit von 32.52%. Diese Ungenauigkeit nimmt mit zunehmender Feinheit der Gitter ab.

Betrachtet man nun die Ergebnisse für die Simulationen mit SGS-Modell fällt auf, dass diese auf den groben Gittern die globale Sherwood-Zahl besser abbilden als die lineare Interpolation ohne SGS-Modell. Für alle vier Modell-Varianten ist die Ungenauigkeit auf allen Gittern kleiner als 4%. Dabei kann es vorkommen, dass für eine Variante mit zunehmender Gitterauflösung die Sherwood-Zahl abwechselnd einmal über- und unterschätzt wird. Die Modelle für Datensatz 2 schneiden insgesamt besser ab als jene für Datensatz 1. Dies gilt jedoch nicht bei dem größten Gitter. Datensatz 2 weist eine durchschnittliche betragsmäßige Abweichung von 0.87% auf, wo hingegen bei Datensatz 1 eine Abweichung von 1.57% vorliegt. Der Vergleich zwischen den Feature-Kombinationen α und β zeigt, dass die Kombination α eine durchschnittliche betragsmäßige Abweichung von 1.36% und Kombination B von 1.08% aufweist. Die Hinzunahme von Feature $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ ist schwierig zu beurteilen. Für Datensatz 1 werden die Abweichungen geringer. Bei Datensatz 2 tritt eine Verschlechterung bei Hinzunahme von $\partial_{\tilde{n}} \tilde{U}_{\tilde{n}}$ als Feature auf. Auf “Mesh 0” hat “Modell 1- β ” die geringste Abweichung. Insgesamt schneidet “Modell 2- α ” am besten ab, mit einer durchschnittlichen Abweichung von 0.68%.

Ein Vergleich mit der Literatur zeigt, dass die hier verwendeten SGS-Modelle für diesen Testfall ähnlich gute Ergebnisse liefern. Vergleicht man Tabelle 22 mit Tabelle 8 aus [12] (“single reaction bei water-air, $d_b = 1.0 \text{ mm}$ ”) zeigt sich, dass “Modell 2- α ” kleinere Abweichungen aufweist.

Der Testfall erlaubt auch eine Variation der Damköhler-Zahl. Hiermit kann die Verallgemeinerungsfähigkeit der SGS-Modelle aus den unterschiedlichen Datensätze beurteilt werden. Für “Set 1” liegt dieser Fall außerhalb des Parameterbereichs, mit dem die Trainingsdaten generiert wurden. Es wird eine geringere Damköhler-Zahl gewählt, $Da = 0.2$. Dieser Fall wird nur mit den SGS-Modellen für die Feature-Kombination α untersucht, da diese bei dem vorherigen Testfall im Ganzen betrachtet besser abgeschnitten haben, als die Version β . In Tabelle 23 sind die Ergebnisse für eine Damköhler-Zahl $Da = 0.2$ aufgelistet.

Simulation	Mesh 0	Mesh 1	Mesh 2	Mesh 3	Ø Abweichung
linear	174.58 (−31.32%)	237.41 (−6.61%)	253.64 (−0.22%)	254.21	
Modell 1- α	260.54 (+2.49%)	263.16 (+3.52%)	263.41 (+3.62%)	254.54 (+0.13%)	2.44 %
Modell 2- α	259.90 (+2.24%)	256.80 (+1.02%)	254.52 (+0.12%)	255.60 (+0.55%)	0.98 %

Tabelle 23: Globale Sherwood-Zahlen für den Testfall mit Damköhler-Zahl $Da = 0.2$ ($Pe = 50000$, $\tilde{c}_{B,0} = 1$).

Auch bei abweichender Damköhler-Zahl zwischen dem untersuchten Fall und den Trainingsdaten für die jeweiligen SGS-Modelle, liefern die SGS-Modelle gute Ergebnisse. Die Abweichungen auf den groben Gittern sind deutlich kleiner verglichen mit der linearen Interpolation ohne SGS-Modell. Insgesamt sind die betragsmäßigen Abweichungen für alle Gitter kleiner als 4%. Auffällig ist, dass die Abweichungen für “Modell 1- α ” mit feiner werdendem Gitter zunehmen. Auf “Modell 2- α ” trifft dies nicht zu. Dies hat im Mittel auch die kleineren Abweichungen. Im Vergleich zum Fall mit $Da = 0.3$ ist die Leistung der SGS-Modelle auf den gröberen Gittern schlechter. Das “Modell 1- α ” zeigt außerhalb seines Parameterbereichs ähnliche Leistungen wie für den Trainingsfall.

Analog für die Variation der Damköhler-Zahl kann auch die Rand-/Anfangsbedingung für die Spezies B $\tilde{c}_{B,0}$ variiert werden. Für beide Datensätze liegt dies innerhalb des Parameterbereichs, mit dem sie trainiert wurden. Für diesen Fall werden ebenfalls jeweils die Version α der Feature-Kombination benutzt. Die Ergebnisse sind in Tabelle 24 zu finden. Es wird $\tilde{c}_{B,0} = 2$ verwendet.

Simulation	Mesh 0	Mesh 1	Mesh 2	Mesh 3	Ø Abweichung
linear	178.87 (−35.98%)	251.84 (−9.86%)	276.62 (−0.99%)	279.38	
Modell 1- α	279.64 (+0.09%)	283.86 (+1.60%)	286.20 (+2.44%)	279.43 (+0.02%)	1.04 %
Modell 2- α	277.75 (−0.58%)	276.32 (−1.10%)	280.21 (+0.30%)	281.35 (+0.71%)	0.67 %

Tabelle 24: Globale Sherwood-Zahlen für den Testfall mit $\tilde{c}_{B,0} = 2$ ($Pe = 50000$, $Da = 0.3$).

Auffällig bei diesen Ergebnissen ist, dass “Modell 1- α ” auf dem größten Gitter sehr kleine Abweichungen hat, diese jedoch mit feiner werdendem Gitter zunehmen. “Modell 2- α ” bildet die Sh -Zahl auf allen Gittern gut ab. Für beide Modelle sind die betragsmäßigen Abweichungen kleiner als 4%.

7 Fazit

Zusammenfassung

In dieser Arbeit wurde eine Möglichkeit behandelt, um Simulationen des Stofftransports um aufsteigende Gasblasen bei großen Schmidt-Zahlen zu beschleunigen. Es wurden mehrere Varianten eines datenbasierten Subgridskalen-Modells entwickelt, sodass gröbere Gitter für die Simulation genutzt werden können und trotzdem nur ein geringer Genauigkeitsverlust bei der Verteilung der Spezies in der Grenzschicht auftritt. Das Subgridskalen-Modell wurde durch die Anwendung eines *Machine Learning* Algorithmus, eines sogenannten neuronalen Netzes, berechnet. Die dafür notwendigen Trainingsdaten wurden durch ein eindimensionales Modell für den Stofftransport generiert. Dieses Modell erlaubt es Trainingsdaten im Rahmen von wenigen Minuten zu berechnen, anstatt eine mehrstündige Simulation durchzuführen. Das eindimensionale Modell wurde durch eine Analyse des Stofftransports einer umströmten Blase auf Basis von Simulationsdaten abgeleitet. Weiterhin wurden Algorithmen zur Reduktion von Datenmengen entwickelt und vorgestellt, sodass nur ein geringer Informationsverlust auftritt. Zwei verschiedene Algorithmen wurden miteinander kombiniert und in den Prozess der Trainingsdatenaufbereitung integriert. Dies hält den Zeitaufwand für die Reduktion möglichst gering. Die Reduktion von redundanten Daten ermöglicht ein schnelleres Training der neuronalen Netze. Es konnten Datenmengen von mehreren Millionen Datenpunkten auf Zehntausende reduziert werden, wodurch die Trainingszeit pro Netz auf wenige Minuten verkürzt wird. Die Validierung der SGS-Modell-Varianten an einem geometrisch einfachen Testfall zeigte eine gute Genauigkeit im Vergleich zu einer Referenzsimulation. Auf allen getesteten Gittern war die Abweichung immer $< 4\%$. Die Abweichungen der besten Variante waren immer geringer als 1.6% . Auch bei Parametervariationen innerhalb des trainierten Datensatzes blieb der Fehler $< 4\%$. Ein Vergleich mit anderen Arbeiten, in denen Subgridskalen-Modelle eingesetzt werden, ist möglich. Im Gegensatz zu den hier an einem einfachen, eindimensionalen Modell gewonnenen Daten wurden in [12] Simulationsdaten zum Trainieren des SGS-Modells genutzt. Trotzdem wird eine ähnliche Genauigkeit erreicht.

Verbesserungspotential/Ausblick

Das Subgridskalen-Modell wurde in dieser Arbeit nur an einem einfachen Fall validiert. Die Blasenform ähnelt der des ursprünglichen Ersatzproblems, dadurch lassen sich schwer Aussagen über die Verallgemeinerungsfähigkeit des Modells treffen. Es ist notwendig das Modell an komplexeren statischen Blasenformen zu testen. Als Beispiel seien für den Anfang sowohl einfache Ellipsoide, die durchaus ein andersartiges Strömungsfeld im Vergleich zur sphärischen Blase haben können (Rezirkulationsgebiete), als auch die in [20] aufgeführten Blasenformen genannt. Tests an Blasen mit veränderlichem Rand und Vergleiche mit experimentellen Daten können folgende Schritte sein. Die Validierung in Kapitel 6 geschah an einem kleinen Ausschnitt des Parameterbereichs. Es muss auch die Leistung des Modells bei anderen Pe -Zahlen und Da -Zahlen untersucht werden. Insbesondere bei kleinen Pe -Zahlen, da dort die Qualität der Trainingsdaten durch die angesprochenen Oszillationen verrauscht ist. Sollte sich bei diesen Untersuchungen zeigen, dass das Modell signifikante Schwächen hat, so gibt es einige Maßnahmen um das Modell zu verbessern. Eine Modellierung des in dieser Arbeit vernachlässigten zweiten Konvektionsterms würde den Einfluss des tangentialen Stofftransports in den Trainingsdaten abbilden. Außerdem kann auch eine Erweiterung des 1D-Modells um den Fall eines Ellipsoids als Ersatzproblem in Betracht gezogen werden. Mit Hilfe von "Oblate spheroidal"-Koordinaten, siehe [21], sollte eine Term-Auswertung wie in Kapitel 3.6 an einem Ellipsoid machbar sein. Die anschließende Modellierung einzelner Terme dürfte jedoch anspruchsvoller werden. Eine Erweiterung hätte den Vorteil auch andere Strömungscharakteristika, wie Rezirkulationsgebiete, in die Trainingsdaten einfließen lassen zu können. Eine weitere Möglichkeit das SGS-Modell zu verbessern ist die Regression der Trainingsdaten an sich zu verbessern. Durch eine Hyperparameter-Optimierung können eventuell geeignetere Netzparameter gefunden werden. Das Round-Verfahren für die Datenreduktion ließe sich verbessern, indem man eine Möglichkeit schafft auf bestimmte Zwischenwerte der Nachkommastellen zu runden. Hier ist jedoch zu prüfen inwieweit dies den Algorithmus verlangsamt.

Symbolverzeichnis

Symbol	Beschreibung	Einheit
c_i	Konzentration Spezies i	mol/m^3
$c_{A,\Sigma}$	Konzentration Spezies A an Grenzfläche Σ	mol/m^3
$c_{B,0}$	Konzentration Spezies B; Anfangs-/Randbedingung; Variationsparameter	mol/m^3
d_b	Durchmesser der Blase	m
d_{eq}	Kugelvolumen-äquivalent-Durchmesser	m
$d_{\partial\Omega}$	Durchmesser des äußeren Randes	m
g_i	Gewicht bzgl. Eintrag i bei der Reduktion mit dem Metrik-Verfahren	—
\mathbf{g}	Fallbeschleunigungvektor	m/s^2
k	Geschwindigkeitskonstante der chem. Reaktion	$\text{m}^3/(\text{s mol})$
$\text{kor}r$	Korrelationskoeffizient	—
n	Koordinate; Normalenrichtung	m
\mathbf{n}	Normalenvektor	m
p	Druck	$\text{kg}/(\text{m s}^2)$
r	Radius; Koordinate	m
r_0	Radius der Blase; Offset zwischen (r, Θ) - und (n, τ) -Koordinaten; $d_{eq}/2$	m
t	Zeit	s
\mathbf{u}	Geschwindigkeitsvektor	m/s
w_{ji}	Gewicht einer Neuronen-Verbindung, siehe Kapitel 2.5	—
A_Σ	Flächeninhalt der Grenzfläche Σ	m^2
D	Diffusionskoeffizient	m^2/s
F_D	durchschnittlicher Fehler, siehe Gleichung (31)	mol/m^4
F_f^D	diffusiver Fluss der Fläche f	mol/s
F_f^A	advektiver Fluss der Fläche f	mol/s
L	maximaler Abstand für die Berechnung des 1D-Modells	m
N_k	Anzahl an Kombinationen	—
N_n	Anzahl an räumlichen Punkten	—
N_t	Anzahl an Zeitschritten	—
R	Reaktionsterm-Label; $R = \overline{\tilde{c}_A \tilde{c}_B}$	—
R_i	Reaktionsterm der Stofftransportgleichung für Spezies i	$\text{mol}/(\text{m}^3 \text{s})$
S_f	Flächeninhalt der Fläche f	m^2
T	Endzeit für die Berechnung mit dem 1D-Modell	s
U_g	Geschwindigkeit der Blase	m/s
U_{in}	Einflussgeschwindigkeit; Randbedingung	m/s
U_r	Geschwindigkeit in radiale Richtung	m/s
U_Θ	Geschwindigkeit in tangentielle Richtung	m/s
U_n	Geschwindigkeit in Normalenrichtung	m/s
U_τ	Geschwindigkeit in tangentielle Richtung	m/s
V	Volumen	m^3
α	Aktivierungsfunktion eines Neurons, siehe Kapitel 2.5	—
δ_d	dynamische Grenzschichtdicke	m
δ_c	Konzentrationsgrenzschichtdicke	m
δn	Gitterweite	m
δt	Zeitschrittweite	s
μ	dynamische Viskosität	$\text{kg}/(\text{m s})$
ν	kinematische Viskosität	m^2/s
ρ	Dichte	kg/m^3
τ	Tangential-Koordinate	m
χ	Seitenverhältnis einer Ellipse	—
$\Delta_{\max} U_\tau(\text{Re})$	berechneter Vergleichswert, siehe Kapitel 3.4	—
Δv_i	Differenz zweier Datenpunkte bzgl. Eintrag i	—
Θ	Polarwinkel	$^\circ$
Σ	Grenzfläche zwischen Blase und Flüssigkeit	—
Ψ	beispielhafte Variable; Platzhalter	—
Ω	Berechnungsgebiet; Flüssigphase	—
$\partial\Omega$	äußerer Rand des Berechnungsgebietes Ω	—



Kennzahlen		
Symbol	Beschreibung	Einheit
Da	Damköhler-Zahl	—
Pe	Peclet-Zahl	—
Re	Reynolds-Zahl	—
Sc	Schmidt-Zahl	—
Sh	Sherwood-Zahl	—

(Formel)-Abkürzungen		
Symbol	Beschreibung	Einheit
loc	lokale Größe	
num	numerisch berechnet (z.B. Simulation)	
int	interpolierter Wert	
$ReLU$	Rectified Linear Unit	
SGS	Subgrid-Scale (deutsch Subgridskalen)	
$\partial_x \Psi$	Ableitung von Ψ nach x	
$\Psi _l$	ausgewertet an der Stelle l	
f	$f = \frac{\delta t}{Pe (\delta \tilde{n})^2}$	—
g	$g = \frac{\delta \tilde{t}}{2 \delta \tilde{n}}$	—
h	$h = Da \delta \tilde{t}$	—
DP	Datenpunkt	—
KV	Kontrollvolumen	—
A	Matrix, siehe Kapitel 3.6	—
B	Matrix, siehe Kapitel 3.6	—
M	Matrix, siehe Kapitel 3.6	—
N	Matrix, siehe Kapitel 3.6	—
O	Matrix, siehe Kapitel 3.6	—
P	Matrix, siehe Kapitel 3.6	—

Super-/Subskripte		
Symbol	Beschreibung	Einheit
$\hat{\Psi}$	charakteristische Größe	
$\tilde{\Psi}$	entdimensionalisierte Größe	
Ψ^+	durch SGS-Modell berechneter Wert	
$\bar{\Psi}$	gemittelter Wert	
Ψ_{01}	auf Wertebereich $[0, 1]$ skaliert	

Literaturverzeichnis

- [1] B. Aboulhasanzadeh, S. Thomas, M. Taeibi-Rahni, G. Tryggvason. *Multiscale computations of mass transfer from buoyant bubbles*. In: *Chemical Engineering Science* 75 (2012), S. 456 – 467.
- [2] G. Besagni, F. Inzoli, T. Ziegenhein. *Two-Phase Bubble Columns: A Comprehensive Review*. In: *ChemEngineering* 2 (2018), S. 13.
- [3] M. Falcone, D. Bothe, H. Marschall. *3D direct numerical simulations of reactive mass transfer from deformable single bubbles: An analysis of mass transfer coefficients and reaction selectivities*. In: *Chemical Engineering Science* 177 (2018), S. 523 – 536.
- [4] R. Kruse, C. Borgelt, C. Braune, F. Klawonn, C. Moewes, M. Steinbrecher. *Computational Intelligence*. Springer Vieweg, 2. Edition, 2015.
- [5] H. Mittag. *Statistik. Eine Einführung mit interaktiven Elementen*. Springer Spektrum, 5. Auflage (2017).
- [6] C. Pesci, A. Weiner, H. Marschall, D. Bothe. *Computational analysis of single rising bubbles influenced by soluble surfactant*. In: *Journal of Fluid Mechanics* 856 (2018), S. 709 – 763.
- [7] M. Schäfer. *Vorlesungsskript zur Vorlesung: Numerische Berechnungsverfahren*. FNB Tu Darmstadt, 2005.
- [8] M. Schäfer. *Computational engineering: introduction to numerical methods*. Springer-Verlag, 1. Edition, 2006.
- [9] A. Weiner, D. Bothe. *Advanced subgrid-scale modeling for convection-dominated species transport at fluid interfaces with application to mass transfer from rising bubbles*. In: *Journal of Computational Physics* 347.1 (2017), S. 261 – 289.
- [10] A. Weiner, D. Hillenbrand, H. Marschall, D. Bothe. *Data-Driven Subgrid-Scale Modeling for Convection-Dominated Concentration Boundary Layers*. In: *Chemical Engineering & Technology* 42.7 (2019), S. 1349 – 1356.
- [11] A. Weiner, J. Timmermann, C. Pesci, J. Grewe, M. Hoffmann, M. Schlüter, D. Bothe. *Experimental and numerical investigation of reactive species transport around a small rising bubble*. In: *Chemical Engineering Science* X 1 (2019), S. 100007.
- [12] Andre Weiner. *Modeling and simulation of convection-dominated species transfer at rising bubble*. 2020.
- [13] <https://idw-online.de/de/news591980>
<https://www.chemietechnik.de/evonik-optimiert-reaktionen-in-blasensaehlen-reaktoren/>
abgerufen jeweils 29.03.2020.
- [14] <https://webbook.nist.gov/chemistry/fluid/>
abgerufen am 29.03.2020.
- [15] <https://scikit-learn.org/stable/modules/ensemble.html#forest>
<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
abgerufen jeweils 29.03.2020.
- [16] <https://www.youtube.com/watch?v=g9c66TUylZ4>
abgerufen am 29.03.2020.
- [17] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
abgerufen am 29.03.2020.
- [18] https://github.com/AndreWeiner/phd_notebooks/blob/master/notebooks/create_sgs_model_assembly.ipynb
https://github.com/AndreWeiner/phd_notebooks/blob/master/notebooks/create_sgs_models.ipynb
abgerufen am 29.03.2020.
- [19] https://github.com/AndreWeiner/phd_notebooks/blob/master/notebooks/sgs_model_evaluation.ipynb
abgerufen am 29.03.2020.
- [20] https://github.com/AndreWeiner/phd_openfoam, abgerufen am 29.03.2020.
- [21] https://en.wikipedia.org/wiki/Oblate_spheroidal_coordinates
abgerufen am 29.03.2020.

Abbildungsverzeichnis

1	Ein Blasensäulenreaktor (Bild 6 (b) in [2]).	5
2	Schematische Darstellungen der dynamischen Grenzschichtdicke δ_d und der Stoffgrenzschichtdicke δ_c um eine mit der Geschwindigkeit U_g aufsteigenden Blase in ruhender Flüssigkeit fl	6
3	Skizze charakteristische Größen.	8
4	Effekt unzureichender Auflösung auf Spezies A (aus [12]).	12
5	Effekt unzureichender Auflösung auf Spezies B (aus [12]).	13
6	Effekt unzureichender Auflösung auf Spezies P (aus [12]).	13
7	Effekt unzureichender Auflösung auf Reaktionsterm (aus [12]).	14
8	Skizze der ersten Zellen an der Grenzfläche (aus [12]).	14
9	Beispielhafte Skizze eines neuronalen Netzes übernommen aus [12].	16
10	Skizze des Berechnungsgebietes mit Randbedingungen.	18
11	Normierte maximale Abweichung $\Delta_{max} U_\tau(Re)$, siehe Gleichung (20).	20
12	Verteilung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche in Abhängigkeit vom Polarwinkel Θ bei $Re = 1$ für unterschiedlich große Berechnungsgebiete. Die gestrichelte Linie ist die Nulllinie.	21
13	Verteilung der tangentialen Geschwindigkeit \tilde{U}_τ an der Grenzfläche in Abhängigkeit vom Polarwinkel Θ bei $Re = 100$ für unterschiedlich große Berechnungsgebiete. Die gestrichelte Linie ist die Nulllinie.	22
14	Gitterstudie mit unterschiedlich feinen Gittern. Verläufe der lokalen Sherwood-Zahl über Polarwinkel Θ (oben), Verläufe der Konzentration der Spezies A in radialer Richtung für $\Theta \approx 92^\circ$ (unten). ($Re = 250$, $Sc = 500$, $Da = 0.3$, $c_{B,0} = 5$)	24
15	Skizze der Koordinatensysteme.	25
16	Verlauf der Terme der linken Seite aus Gleichung (24) für unterschiedliche Spezies bei $\Theta = 90^\circ$. Oben Spezies A, mitte Spezies B, unten Spezies P ($Pe = 125000$; $Da = 0.3$; $c_{B,0} = 5$).	27
17	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).	28
18	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A zu unterschiedlichen Zeiten bei Polarwinkel $\Theta = 90^\circ$. Oben $\tilde{t} = 5$, Mitte $\tilde{t} = 0.375$, unten $\tilde{t} = 0.125$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).	29
19	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies A bei unterschiedlichen Péclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$).	30
20	Verlauf der Geschwindigkeit \tilde{U}_n in Normalenrichtung für unterschiedliche Polarwinkel Θ ($Pe = 125000$).	31
21	Verlauf der Tangentialgeschwindigkeit \tilde{U}_τ über dem Polarwinkel Θ für unterschiedliche Abstände \tilde{n} ($Pe = 125000$).	31
22	Verlauf von $\frac{\partial c_A}{\partial \tilde{r}}$ in Normalenrichtung für verschiedene Polarwinkel Θ (oben). Verlauf von $\frac{\partial c_A}{\partial \tilde{r}}$ in tangentialer Richtung für verschiedene Abstände \tilde{n} (unten). ($Pe = 125000$; $Da = 0.3$; $c_{B,0} = 5$).	32
23	Verlauf von $\partial_{\tilde{n}} \tilde{U}_n$ an der Grenzfläche über dem Polarwinkel Θ für verschiedene Reynolds-Zahlen.	33
24	Auftretende Oszillationen in der Crank-Nicolson-Lösung, bei $N_n = 421$ und $N_t = 10001$	36
25	Auftreten von Oszillationen für $Pe = 5000$ (links), $Pe = 60000$ (Mitte), $Pe = 125000$ (rechts). Auf der x-Achse sind N_n und auf der y-Achse die N_t aufgetragen. Die Zahlen verdeutlichen bis zu welchem Zeitschritt Oszillationen auftreten.	36
26	Vergleich zwischen Mittelwert-Verlauf und oszillierendem Verlauf bei $Pe = 5000$ und $N_n = 421$	37
27	Einfluss der Zeitschritt-Anzahl bei gleichbleibendem N_n auf die Oszillation. $Pe = 60000$, $Da = 0.15$, $N_n = 421$	37
28	Einfluss der Gitterpunkt-Anzahl bei gleichbleibendem N_t . $Pe = 60000$, $Da = 0.15$, $N_n = 10001$	38
29	Vergleich der zufallsgenerierten Daten (schwarz, oben) mit den durch Verfahren Round, eine Nachkommastelle, (rot, Mitte) und Verfahren Metrik, Schwellenwert 0.01, (blau, unten) reduzierten Daten.	42
30	Vergleich der mit dem Verfahren Metrik reduzierten Daten mit unterschiedlichen Schwellenwerten: schwarz 0.1 (oben); rot 0.05 (Mitte); blau 0.01 (unten).	43
31	Skizze eines beispielhaften Decision Tree.	46
32	Bravais-Pearson Korrelationskoeffizienten zwischen Features (links), Features und Label (rechts) des Datensatz 1.	48
33	“Feature Importance” des Random Forest Regressors für die Label des Datensatz 1.	48
34	Bravais-Pearson Korrelationskoeffizienten zwischen Features (links), Features und Label (rechts) des Datensatz 2.	49
35	“Feature Importance” des Random Forest Regressors für die Label des Datensatz 2.	50
36	Histogramm für Abweichungen der Label-Werte des Trainingsdatensatzes gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies A.	52
37	Histogramm für Abweichungen der Label-Werte des Testdatensatzes, “Set 1- α ”, gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies A.	52

38	Histogramm für Abweichungen der Label-Werte des Trainingsdatensatzes gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies B und P	53
39	Histogramm für Abweichungen der Label-Werte des Testdatensatzes, “Set 1- α ”, gegenüber den von “Modell 2- α ” berechneten Label-Werten $^+$ für die Label der Spezies B und P	53
40	Histogramm für Abweichungen der Label-Werte der Datensätze gegenüber den von “Modell 2- α ” berechneten Label-Werten R^+ für das Label Reaktionsterm. Links für “Set 2- α , rechts “Set 1- α	54
41	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies B bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).	64
42	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies P bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).	65
43	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies B bei unterschiedlichen Peclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$). . . .	66
44	Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies P bei unterschiedlichen Peclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$). . . .	67
45	Grafische Darstellung der Arbeitsschritte des Round-Verfahrens an einer beispielhaften Datenpunktmenge. .	69
46	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für die Label der Spezies A.	70
47	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für die Label der Spezies B und P	70
48	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für das Reaktionsterm-Label.	71
49	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- β ” für die Label der Spezies A.	71
50	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- β ” für die Label der Spezies B und P	71
51	Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- β ” für das Reaktionsterm-Label.	72
52	Histogramme bezüglich des Trainingsdatensatzes für “Modell 2- β ” für die Label der Spezies A.	72
53	Histogramme bezüglich des Trainingsdatensatzes für “Modell 2- β ” für die Label der Spezies B und P	72
54	Histogramme bezüglich des Trainingsdatensatzes für “Modell 2- β ” für das Reaktionsterm-Label.	73

Tabellenverzeichnis

1	Entdimensionalisierungsvorschrift.	9
2	Dimensionslose Kennzahlen.	10
3	Für die Korrektur benötigte Größen bei Benutzung des SGS-Modells.	15
4	OpenFOAM Randbedingungen für Geschwindigkeit, Druck, Spezies A, B und P	19
5	Anfangsbedingungen für Geschwindigkeit, Druck, Spezies A, B und P	19
6	Charakteristische Größen.	19
7	Zu definierende Größen.	20
8	Abweichungen für $\frac{d\Omega}{d\epsilon} = 7$	21
9	Globale Sherwood-Zahl.	23
10	Erläuterung der Legendenbezeichnungen in den Abbildungen 16, 17, 18 und 19.	26
11	Randbedingungen für eindimensionales Modell.	34
12	Abkürzungen	34
13	Anfangsbedingungen für eindimensionales Modell.	35
14	Durchschnittliche Fehler F_D für $N_t = 10001$ im Vergleich zu $N_n = 421$, $N_t = 75001$	38
15	Vergleich der Reduktionsverfahren und Schwellenwerte	41
16	Vergleich der Reduktionsverfahren an realistischen Daten.	44
17	Kombination beider Reduktionsverfahren.	44
18	Parameterbereiche der unterschiedlichen Datensätze.	45
19	Verwendete Hyperparameter für den <i>Random Forest Regressor</i>	47
20	Verwendete Feature Kombinationen für Datensatz 1 und 2.	49
21	Datensatzgröße und Trainingszeiten für die unterschiedlichen Datensatz-Versionen.	51
22	Globale Sherwood-Zahlen für die unterschiedlichen Simulationen und unterschiedliche Gitter.	55
23	Globale Sherwood-Zahlen für den Testfall mit Damköhler-Zahl $Da = 0.2$ ($Pe = 50000$, $\tilde{c}_{B,0} = 1$).	56
24	Globale Sherwood-Zahlen für den Testfall mit $\tilde{c}_{B,0} = 2$ ($Pe = 50000$, $Da = 0.3$).	56
25	Abkürzungen	68

Anhang

A Herleitung des (n, τ) -Koordinatensystems

Abbildung 15 zeigt den skizzierten Zusammenhang zwischen den Kugelkoordinaten (r, Θ, ϕ) und dem lokalen Koordinatensystem (n, τ, ϕ) . Aufgrund der Rotationssymmetrie wird die Umfangsrichtung ϕ bei beiden Koordinatensystemen weggelassen.

Aus Abbildung 15 lassen sich folgende Zusammenhänge ablesen, siehe Gleichung (36). Mit \mathbf{e} wird der jeweilige Einheitsvektor bezeichnet.

$$\begin{aligned}\mathbf{e}_r &= \mathbf{e}_n \\ \mathbf{e}_\Theta &= \mathbf{e}_\tau\end{aligned}\tag{36}$$

Der Gradient in Kugelkoordinaten ergibt sich zu Gleichung (37).

$$\nabla = \frac{\partial}{\partial r} \mathbf{e}_r + \frac{1}{r} \frac{\partial}{\partial \Theta} \mathbf{e}_\Theta\tag{37}$$

Damit ergeben sich

$$\begin{aligned}U_r &= U_n \\ U_\Theta &= U_\tau\end{aligned}\tag{38}$$

und

$$\begin{aligned}\frac{\partial}{\partial r} &= \frac{\partial}{\partial n} \\ \frac{1}{r} \frac{\partial}{\partial \Theta} &= \frac{\partial}{\partial \tau}\end{aligned}\tag{39}$$

woraus

$$\begin{aligned}\frac{\partial^2}{\partial r^2} &= \frac{\partial^2}{\partial n^2} \\ \frac{1}{r^2} \frac{\partial^2}{\partial \Theta^2} &= \frac{1}{r} \frac{1}{r} \frac{\partial^2}{\partial \Theta^2} = \frac{\partial^2}{\partial \tau^2} = \frac{\partial^2}{\partial \tau^2}\end{aligned}\tag{40}$$

folgt. Außerdem gilt $r = r_0 + n$.

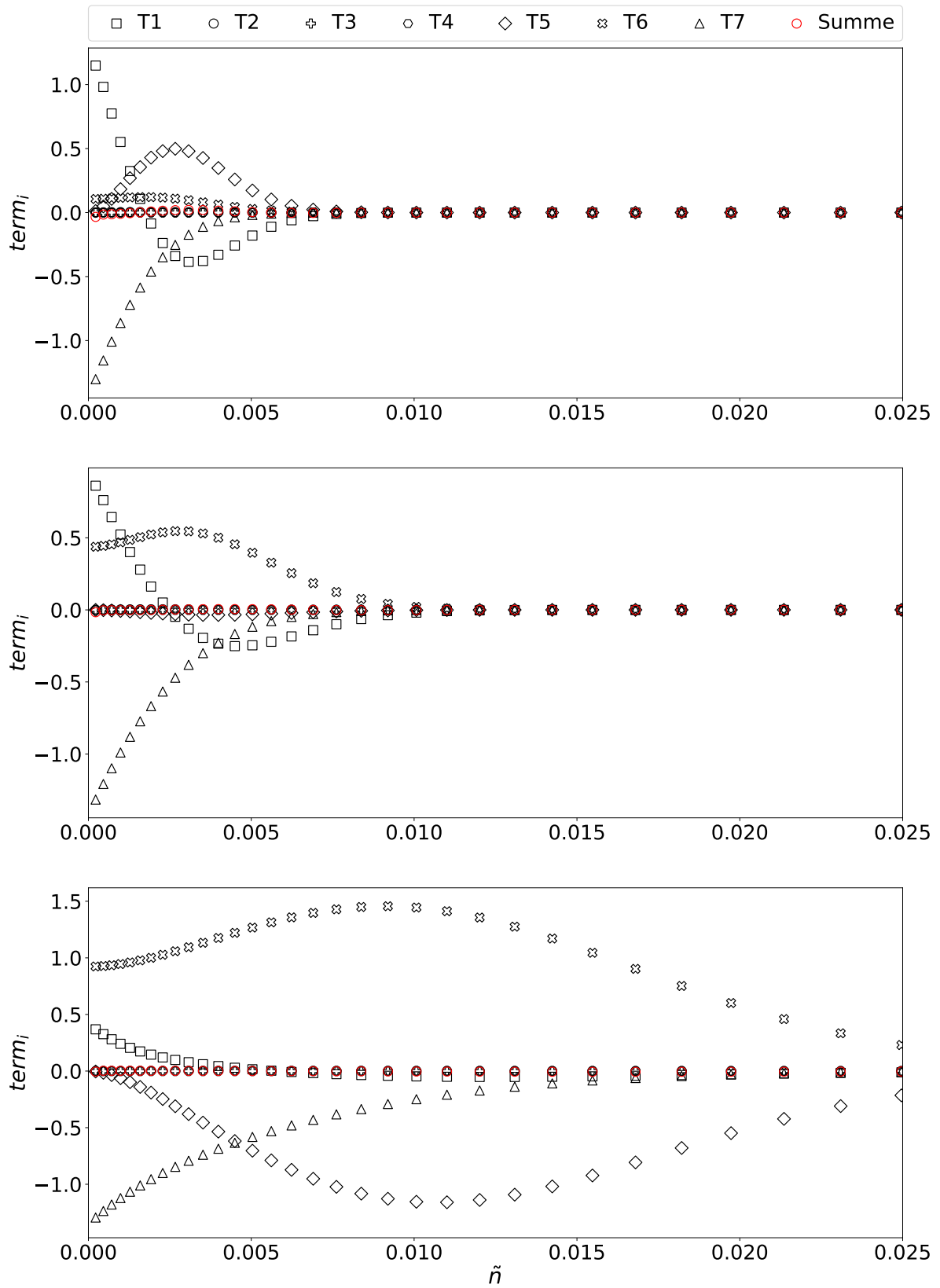


Abbildung 41: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies B bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).

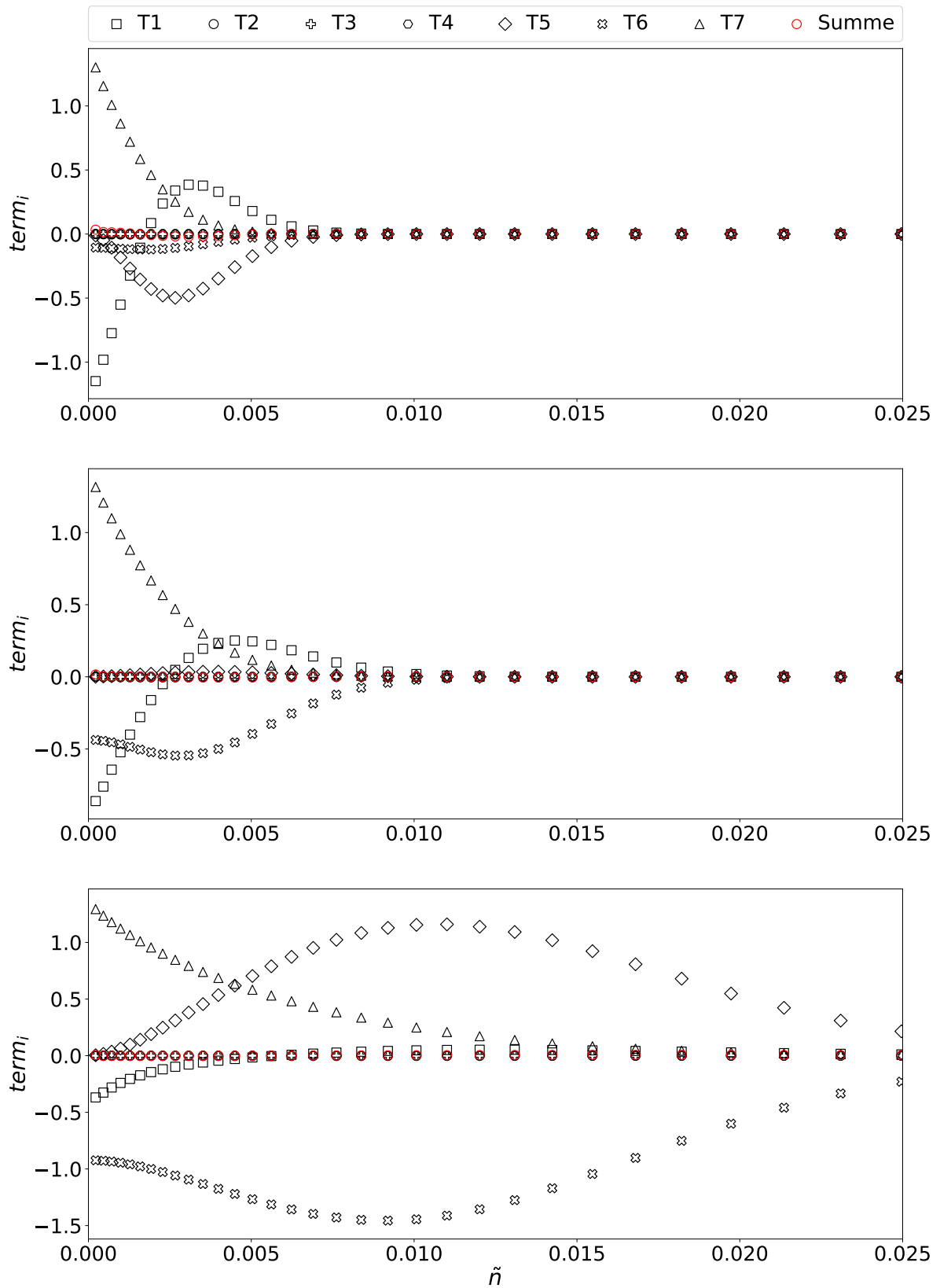


Abbildung 42: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies P bei unterschiedlichen Polarwinkeln Θ . Oben $\Theta = 45^\circ$, mitte $\Theta = 90^\circ$, unten $\Theta = 135^\circ$ ($Pe = 50000$; $Da = 0.3$; $c_{B,0} = 5$).

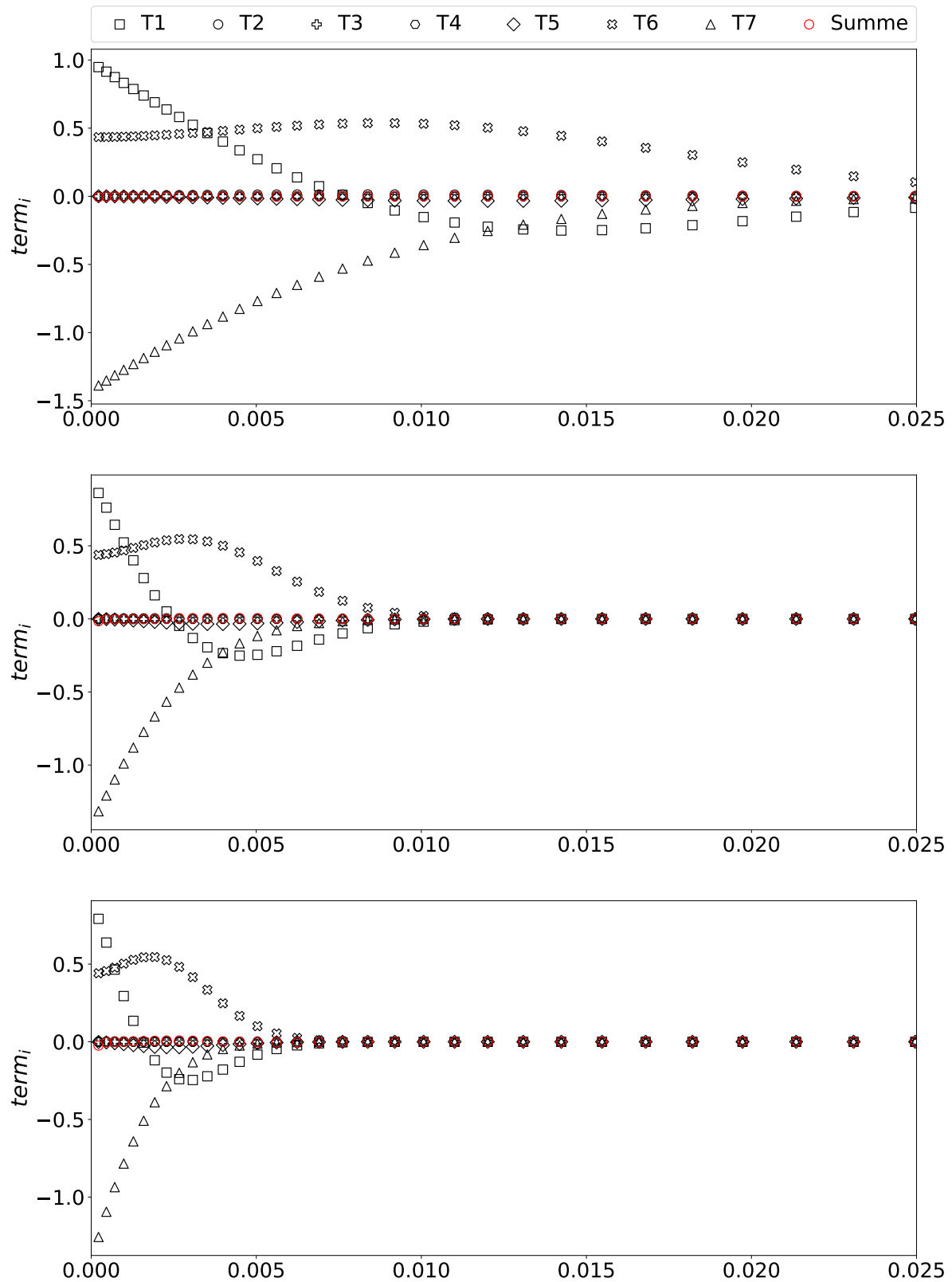


Abbildung 43: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies B bei unterschiedlichen Peclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$).

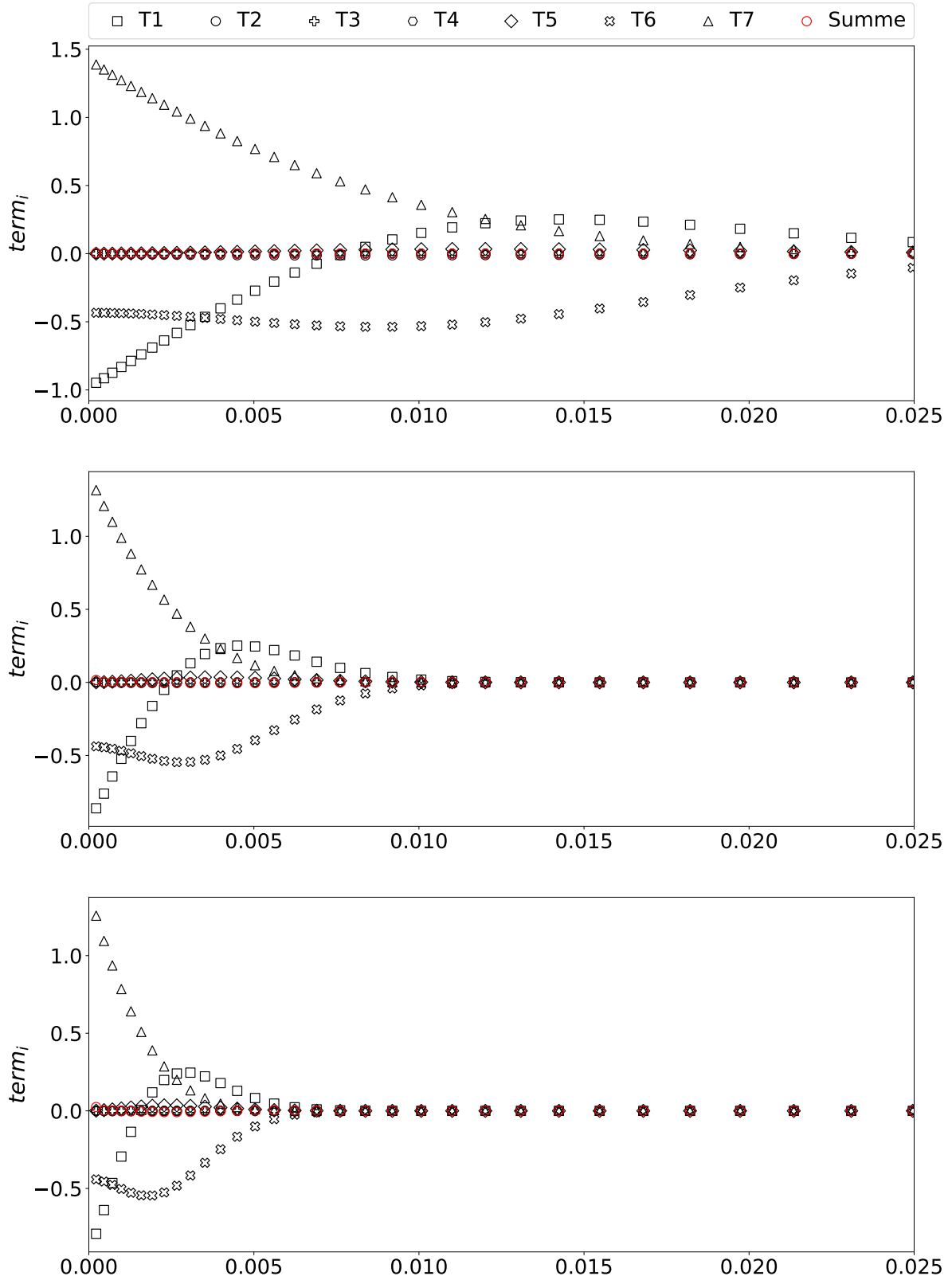


Abbildung 44: Verlauf der Terme der linken Seite aus Gleichung (24) für Spezies P bei unterschiedlichen Peclet-Zahlen bei Polarwinkel $\Theta = 90^\circ$. Oben $Pe = 5000$, mitte $Pe = 50000$, unten $Pe = 125000$ ($Da = 0.3$; $c_{B,0} = 5$).

C Crank-Nicolson-Diskretisierung

Die Diskretisierungsvorschriften nach dem Crank-Nicolson-Verfahren lauten:

$$\begin{aligned}
 \frac{\partial c}{\partial t} &= \frac{c_j^{i+1} - c_j^i}{\delta t} \\
 \frac{\partial c}{\partial n} &= \frac{1}{4 \delta n} (c_{j+1}^{i+1} - c_{j-1}^{i+1} + c_{j+1}^i - c_{j-1}^i) \\
 \frac{\partial^2 c}{\partial n^2} &= \frac{1}{2 \delta n^2} (c_{j+1}^{i+1} - 2c_j^{i+1} + c_{j-1}^{i+1} + c_{j+1}^i - 2c_j^i + c_{j-1}^i) \\
 c &= \frac{1}{2} (c_j^{i+1} + c_j^i)
 \end{aligned} \tag{41}$$

D Gleichungssysteme 1D-Modell

Analog zu Gleichung (29) werden hier die in Gleichung (30) vorkommenden Matrizen B , N , P und O detailliert dargestellt.

Eintrag	Term
$a_{1,j}$	$\tilde{U}_{\bar{n},j} \frac{g}{2} - \frac{f}{2}$
$a_{2,j}$	$1 + f + \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i)$
$a_{3,j}$	$-\tilde{U}_{\bar{n},j} \frac{g}{2} - \frac{f}{2}$
$a_{4,j}$	$1 - f - \frac{h}{4} (\tilde{c}_{B,j}^{i+1} + \tilde{c}_{B,j}^i)$
$b_{2,j}$	$1 + f + \frac{h}{4} (\tilde{c}_{A,j}^{i+1} + \tilde{c}_{A,j}^i)$
$b_{4,j}$	$1 - f - \frac{h}{4} (\tilde{c}_{A,j}^{i+1} + \tilde{c}_{A,j}^i)$
$p_{2,j}$	$1 + f$
$p_{4,j}$	$1 - f$

Tabelle 25: Abkürzungen

$$B = \begin{bmatrix} b_{2,0} & -f & 0 & \cdots & \cdots & 0 \\ a_{3,1} & b_{2,1} & a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_{3,N_n-1} & b_{2,N_n-1} & a_{1,N_n-1} \\ 0 & \cdots & \cdots & 0 & -f & b_{2,N_n} \end{bmatrix} \tag{42}$$

$$N = \begin{bmatrix} b_{4,0} & f & 0 & \cdots & \cdots & 0 \\ -a_{3,1} & b_{4,1} & -a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -a_{3,N_n-1} & b_{4,N_n-1} & -a_{1,N_n-1} \\ 0 & \cdots & \cdots & 0 & f & b_{4,N_n} \end{bmatrix} \tag{43}$$

$$P = \begin{bmatrix} p_{2,0} & -f & 0 & \dots & \dots & 0 \\ a_{3,1} & p_{2,1} & a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_{3,N_n-1} & p_{2,N_n-1} & a_{1,N_n-1} \\ 0 & \dots & \dots & 0 & -f & p_{2,N_n} \end{bmatrix} \quad (44)$$

$$O = \begin{bmatrix} p_{4,0} & f & 0 & \dots & \dots & 0 \\ -a_{3,1} & p_{4,1} & -a_{1,1} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -a_{3,N_n-1} & p_{4,N_n-1} & -a_{1,N_n-1} \\ 0 & \dots & \dots & 0 & f & p_{4,N_n} \end{bmatrix} \quad (45)$$

E Datenreduktion

Arbeitsschritte Round-Verfahren

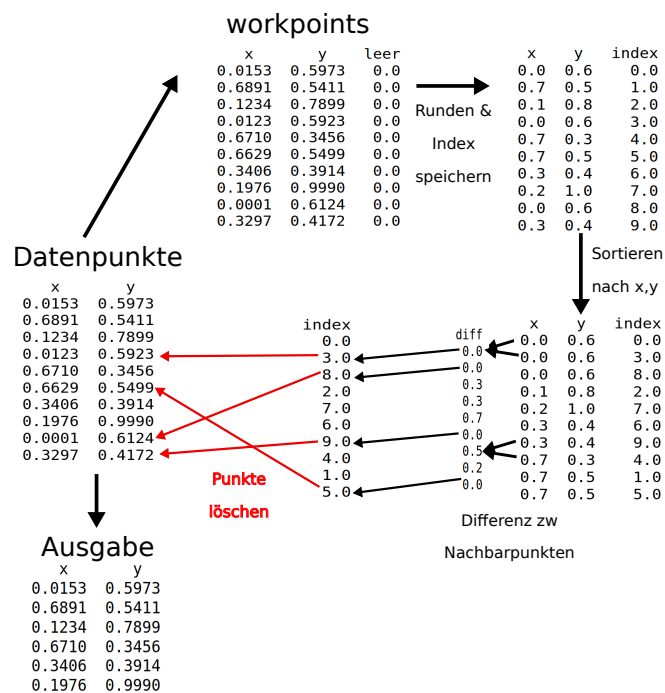


Abbildung 45: Grafische Darstellung der Arbeitsschritte des Round-Verfahrens an einer beispielhaften Datenpunktmenge.

Histogramme für die Abweichungen zwischen Trainingsdaten und vom SGS-Modell berechneten Werten, vgl. Kapitel 5.3.

“Modell 1- α ”

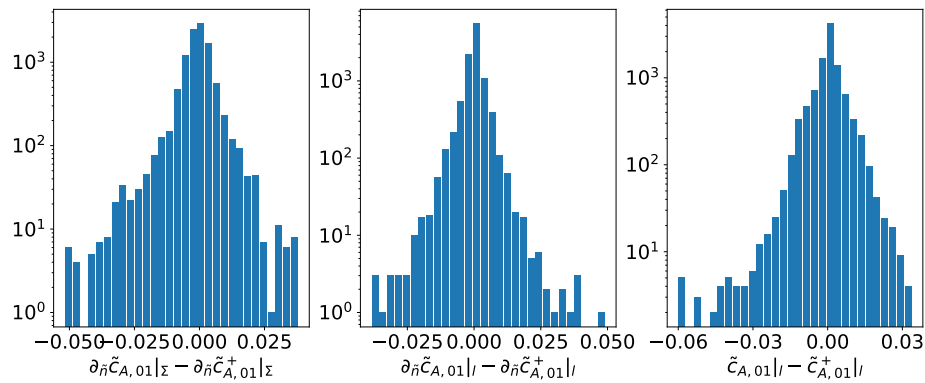


Abbildung 46: Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für die Label der Spezies A.

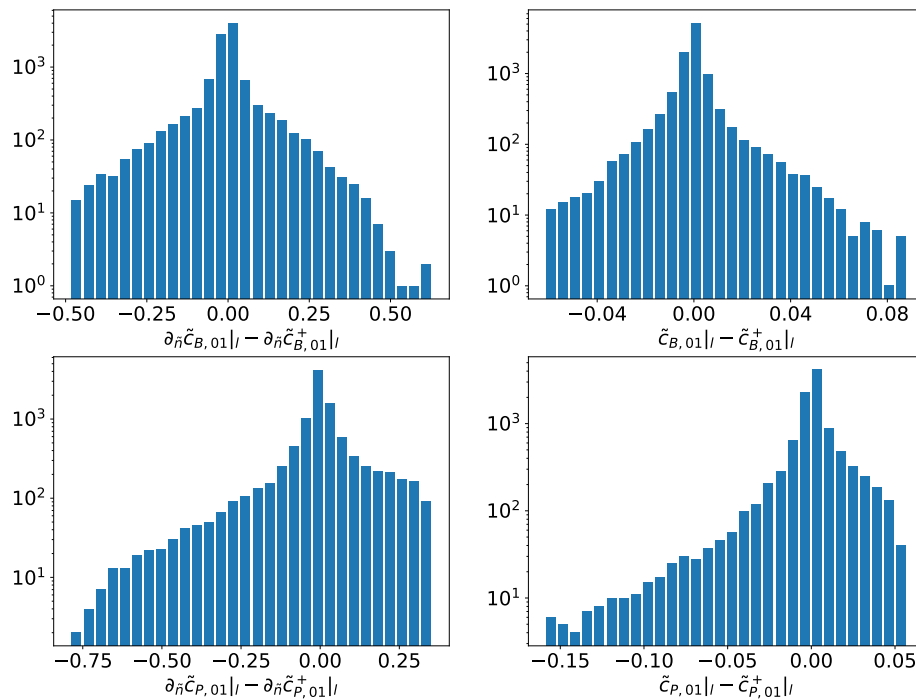


Abbildung 47: Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für die Label der Spezies B und P.

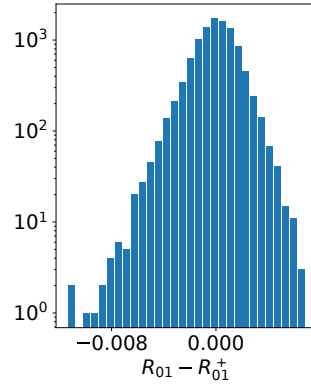


Abbildung 48: Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- α ” für das Reaktionsterm-Label.

“Modell 1- β ”

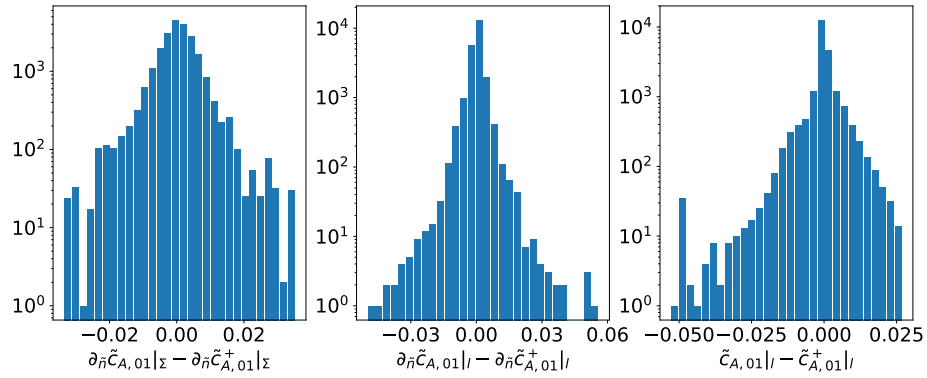


Abbildung 49: Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- β ” für die Label der Spezies A.

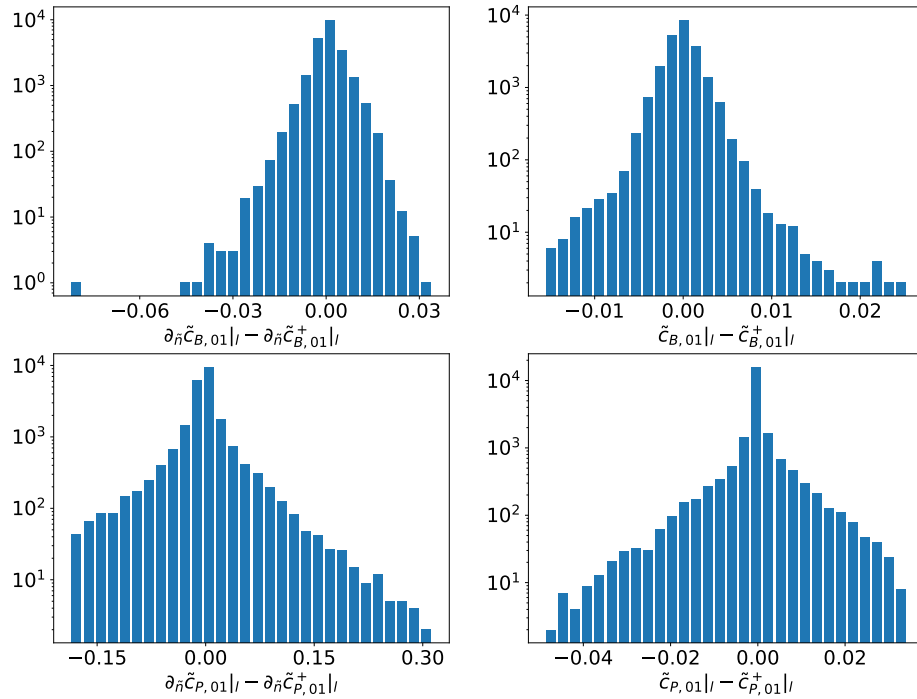


Abbildung 50: Histogramme bezüglich des Trainingsdatensatzes für “Modell 1- β ” für die Label der Spezies B und P.

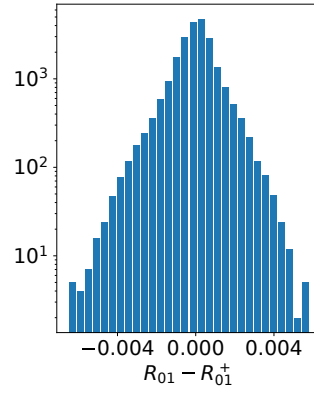


Abbildung 51: Histogramme bezüglich des Trainingsdatensatzes für "Modell 1-β" für das Reaktionsterm-Label.

"Modell 2-β"

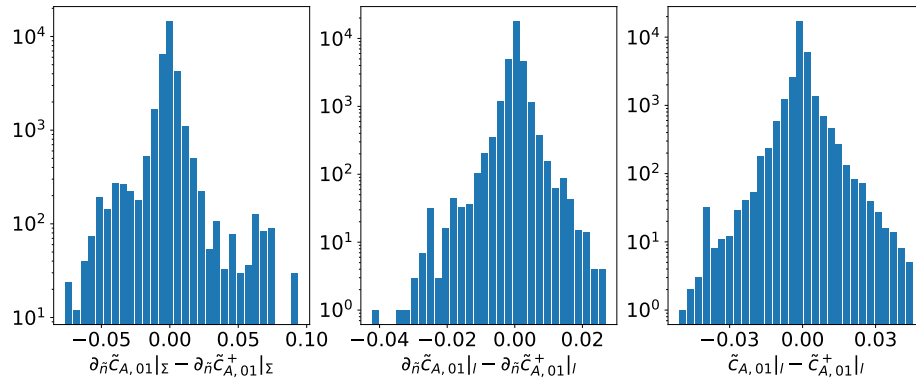


Abbildung 52: Histogramme bezüglich des Trainingsdatensatzes für "Modell 2-β" für die Label der Spezies A.

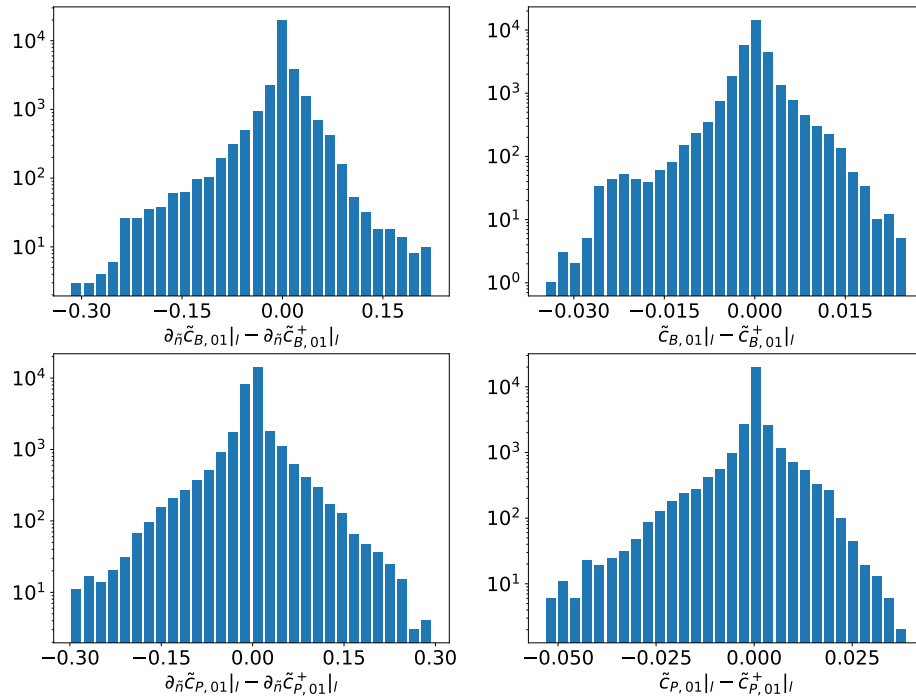


Abbildung 53: Histogramme bezüglich des Trainingsdatensatzes für "Modell 2-β" für die Label der Spezies B und P.

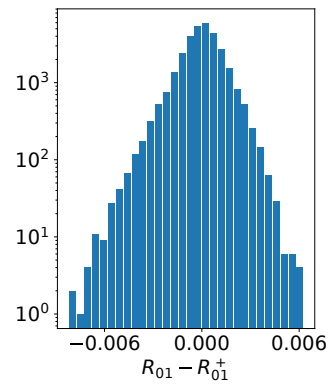


Abbildung 54: Histogramme bezüglich des Trainingsdatensatzes für "Modell 2- β " für das Reaktionsterm-Label.